Internal Regret with Partial Monitoring; Calibration-Based Optimal Algorithms

Vianney Perchet

Internal Regret with Partial Monitoring Calibration-Based Optimal Algorithms

Vianney Perchet*

February 18, 2011

Abstract

We provide consistent random algorithms for sequential decision under partial monitoring, *i.e.* when the decision maker does not observe the outcomes but receives instead random feedback signals. Those algorithms have no internal regret in the sense that, on the set of stages where the decision maker chose his action according to a given law, the average payoff could not have been improved in average by using any other fixed law.

They are based on a generalization of calibration, no longer defined in terms of a Voronoï diagram but instead of a Laguerre diagram (a more general concept). This allows us to bound, for the first time in this general framework, the expected average internal – as well as the usual external – regret at stage n by $O(n^{-1/3})$, which is known to be optimal.

Key Words : Repeated games, On-line learning, Regret, Partial Monitoring, Calibration, Voronoï and Laguerre Diagrams

Hannan [17] introduced the notion of regret in repeated games: a player (that will be referred as a decision maker or also a forecaster) has no external regret if, asymptotically, his average payoff could not have been greater if he had known, before the beginning of the game, the empirical distribution of moves of the other player. Blackwell [6] showed that the existence of such *externally consistent* strategies, first proved by [17], is a consequence of his approachability theorem. A generalization of this result and a more precise notion of regret are due to Foster & Vohra [13] and Fudenberg &

*Centre de Mathématiques et de Leurs Applications UMR 8536, École Normale Supérieure, 61, avenue du président Wilson, 94235 Cachan, France. vianney.perchet@normalesup.

Levine [16]: there exist internally consistent strategies, *i.e.* such that for any of his action, the decision maker has no external regret on the set of stages where he actually chose this specific action. Hart & Mas-Colell [18] also used Blackwell's approachability theorem to construct explicit algorithms that bound the internal (and therefore the external) regret at stage n by $O\left(n^{-1/2}\right)$.

Some of those results have been extended to the partial monitoring framework, *i.e.* where the decision maker receives at each stage a random signal, whose law might depend on his unobserved payoff. Rustichini [27] defined - and proved the existence of - externally consistent strategies, *i.e.* such that the average payoff of the decision maker could not have been asymptotically greater if he had known, before the beginning of the game, the empirical distribution of signals. Actually, the relevant information is a vector of probability distributions, one for each action of the decision maker, that is called *a flag*.

Some algorithms bounding optimally the expected regret by $O\left(n^{-1/3}\right)$ have been exhibited under some strong assumptions on the signalling structure – see Cesa-Bianchi & Lugosi [9], Theorem 6.7 for the optimality of this bound. For example, Jaksch, Ortner & Auer [20] considered the Markov decision process framework, Cesa-Bianchi, Lugosi & Stoltz [10] assumed that payoffs can be deduced from flags and Lugosi, Mannor & Stoltz [23] that feedbacks are deterministic (along with the fact that the worst compatible payoff is linear with respect to the flag). When no such assumption is made, Lugosi, Mannor & Stoltz [23] provided an algorithm (based on the exponential weight algorithm) that bounds regret by $O\left(n^{-1/5}\right)$.

In this framework, internal regret was defined by Lehrer & Solan [21]; stages are no longer distinguished as a function of the action chosen by the decision maker (as in the full monitoring case) but as a function of its law. Indeed, the evaluation of the payoff (usually called *worst case*) is not linear with respect to the flag. So a best response - in a sense to be defined - to a given flag might consist only in a mixed action (*i.e.* a probability distribution over the set of actions). Lehrer & Solan [21] also proved the existence and constructed internally consistent strategies, using the characterization of approachable convex sets due to Blackwell [5]. Perchet [24] provided an alternative algorithm, recalled in section 2.2; this latter is based on calibration, a notion introduced by Dawid [12]. Roughly speaking, these algorithms ε-discretize arbitrarily the space of flags and each point of the discretization is called a possible prediction. Then, stage after stage, they predict what will be the next flag and output a *best response* to it. If the sequence of

predictions is calibrated then the average flag, on the set of stages where a specific prediction is made, will be close to this prediction.

Thanks to the continuity of payoff and signaling functions, both algorithms bound the internal regret by $\varepsilon + O\left(n^{-1/2}\right)$. However the first drawback lies in their computational complexities: at each stage, the algorithm of Perchet [24] solves a system of linear equations while the one Lehrer & Solan [21], after a projection on a convex set, solves a linear program. In both case, the size of the linear system or program considered is polynomial in ε and exponential in the numbers of actions and signals. The second drawback is that the constants in the rate of convergence depend drastically on ε.

As a consequence, a classic *doubling trick* argument will generate an algorithm with a strongly sub-optimal rate of convergence – that might even depend on the size of the actions sets – and a complexity that increases with time.

Our main result is Theorem 2.10, stated in section 2.3: it provides the first algorithm that bounds optimally both internal and external regret by $O\left(n^{-1/3}\right)$ in the general case. It is a modification of the algorithm of Perchet [24] that does not use an arbitrary discretization but constructs carefully a specific one and then computes, stage by stage, the solution of a system of linear equations of constant size. In section 3.1, an other algorithm – based on Blackwell's approachability as the one of Lehrer & Solan [21] – with optimal rate and smaller constants is exhibited; it requires however to solve, at each stage, a linear program of constant size.

Section 1 is devoted to the simpler framework of full monitoring. We recall definitions of calibration and regret and we provide a naïve algorithm to construct strategies with internal regret asymptotically smaller than ε. We show how to modify this algorithm – however in a not efficient way – in order to bound optimally the regret by $O\left(n^{-1/2}\right)$. This has to be seen only as a tool that can be easily adapted with partial monitoring in order to reach the optimal bound of $O\left(n^{-1/3}\right)$; this is done in section 2. Some extensions (the second algorithm, the so-called *compact case* and variants to strengthen the constants) are presented in section 3. Some technical proofs can be found in Appendix.

1 Full monitoring

1.1 Model and definitions

Consider a two-person game Γ repeated in discrete time, where at stage $n \in \mathbb{N}$, a decision maker, or forecaster, (resp. the environment or Nature) chooses an action $i_n \in \mathcal{I}$ (resp. $j_n \in \mathcal{J}$). This generates a payoff $\rho_n = \rho(i_n, j_n)$, where ρ is a mapping from $\mathcal{I} \times \mathcal{J}$ to $\mathbb{R}$, and a regret $r_n \in \mathbb{R}^I$ defined by:

$$r_n = \Big[\rho(i, j_n) - \rho(i_n, j_n)\Big]_{i \in \mathcal{I}} \in \mathbb{R}^I,$$

where I is the finite cardinality of $\mathcal{I}$ (and J the one of $\mathcal{J}$). This vector represents the differences between what the decision maker could have got and what he actually got.

The choices of i_n and j_n depend on the past observations (also called finite history) $h_{n-1} = (i_1, j_1, \ldots, i_{n-1}, j_{n-1})$ and may be random. Explicitly, the set of finite histories is denoted by $H = \bigcup_{n \in \mathbb{N}} (\mathcal{I} \times \mathcal{J})^n$, with $(\mathcal{I} \times \mathcal{J})^0 = \emptyset$ and a strategy σ of the decision maker is a mapping from H to $\Delta(\mathcal{I})$, the set of probability distributions over $\mathcal{I}$. Given the history $h_n \in (\mathcal{I} \times \mathcal{J})^n$, $\sigma(h_n) \in \Delta(\mathcal{I})$ is the law of i_{n+1}. A strategy τ of Nature is defined similarly as a function from H to $\Delta(\mathcal{J})$. A pair of strategies (σ, τ) generates a probability, denoted by $\mathbb{P}_{\sigma,\tau}$, over $(\mathcal{H}, \mathcal{A})$ where $\mathcal{H} = (\mathcal{I} \times \mathcal{J})^{\mathbb{N}}$ is the set of infinite histories embedded with the cylinder σ-field.

We extend the payoff mapping ρ to $\Delta(\mathcal{I}) \times \Delta(\mathcal{J})$ by $\rho(x, y) = \mathbb{E}_{x,y}[\rho(i, j)]$ and for any sequence $a = (a_m)_{m \in \mathbb{N}}$ and any $n \in \mathbb{N}_*$, we denote by $\bar{a}_n = \frac{1}{n} \sum_{m=1}^{n} a_m$ the average of a up to stage n.

Definition 1.1 (Hannan [17]) *A strategy σ of the forecaster is externally consistent if for every strategy τ of Nature:*

$$\limsup_{n \to \infty} \bar{r}_n^i \leq 0, \quad \forall i \in \mathcal{I}, \quad \mathbb{P}_{\sigma,\tau}-\text{as}.$$

In words, a strategy σ is externally consistent if the forecaster could not have had a greater payoff if he had known, before the beginning of the game, the empirical distribution of actions of Nature. Indeed, the external consistency of σ is equivalent to the fact that :

$$\limsup_{n \to \infty} \max_{x \in \Delta(\mathcal{I})} \rho(x, \bar{j}_n) - \bar{\rho}_n \leq 0, \quad \mathbb{P}_{\sigma,\tau}-\text{as}. \tag{1}$$

Foster & Vohra [13] (see also Fudenberg & Levine [16]) defined a more precise notion of regret. The internal regret of the stage n, denoted by

$R_n \in \mathbb{R}^{I\times I}$, is also generated by the choices of i_n and j_n and its (i,k)-th coordinate is defined by:

$$R_n^{ik} = \begin{cases} \rho(k, j_n) - \rho(i, j_n) & \text{if } i = i_n \\ 0 & \text{otherwise.} \end{cases}$$

Stated differently, every row of the matrix R_n is null except the i_n-th which is r_n.

Definition 1.2 (Foster & Vohra [13]) *A strategy σ of the forecaster is internally consistent if for every strategy τ of Nature:*

$$\limsup_{n\to\infty} \bar{R}_n^{ik} \leq 0 \quad \forall i,k \in \mathcal{I}, \quad \mathbb{P}_{\sigma,\tau}-\text{as.}$$

We introduce the following notations to define ε-internally consistency. Denote by $N_n(i)$ the set of stages before the n-th where the forecaster chose action i and $\bar{\jmath}_n(i) \in \Delta(\mathcal{J})$ the empirical distribution of Nature's actions on this set. Formally,

$$N_n(i) = \{m \in \{1,\dots,n\};\ i_m = i\} \quad \text{and} \quad \bar{\jmath}_n(i) = \frac{\sum_{m\in N_n(i)} j_m}{|N_n(i)|} \in \Delta(\mathcal{J}). \tag{2}$$

A strategy is ε-internally consistent if for every $i,k \in \mathcal{I}$

$$\limsup_{n\to\infty} \frac{|N_n(i)|}{n}\Big(\rho\big(k,\bar{\jmath}_n(i)\big) - \rho\big(i,\bar{\jmath}_n(i)\big) - \varepsilon\Big) \leq 0, \quad \mathbb{P}_{\sigma,\tau}-\text{as.}$$

If we define, for every $\varepsilon \geq 0$, the ε-best response correspondence by :

$$BR_\varepsilon(y) = \left\{ x \in \Delta(\mathcal{I});\ \rho(x,y) \geq \max_{z\in\Delta(\mathcal{I})} \rho(z,y) - \varepsilon \right\},$$

then a strategy of the decision maker is ε-internally consistent if any action i is either an ε-best response to the empirical distribution of Nature's actions on $N_n(i)$ or the frequency of i is very small. We will simply denote BR_0 by BR and call it the best response correspondence.

From now on, given two sequences $\big\{l_m \in \mathcal{L}, a_m \in \mathbb{R}^d;\ m \in \mathbb{N}\big\}$ where $\mathcal{L}$ is a finite set, we will define the subset of integers $N_n(l)$ and the average $\bar{a}_n(l)$ as in equation (2).

Proposition 1.3 (Foster & Vohra [13]) *For every $\varepsilon \geq 0$, there exist ε-internally consistent strategies.*

Although the notion of internal regret is a refinement of the notion of external regret (in the sense that any internally consistent strategy is also externally consistent), Blum & Mansour [7] proved that any externally consistent algorithm can be efficiently transformed into an internally consistent one (actually they obtained an even stronger property called *swap consistency*).

Foster & Vohra [13] and Hart & Mas-Colell [18] proved directly the existence of 0-internally consistent strategies using different algorithms (with optimal rates and based respectively on the Expected Brier Score and Blackwell's approachability theorem). In some sense, we merge these two last proofs in order to provide a new one — given in the following section — that can be extended quite easily to the partial monitoring framework.

1.2 A naïve algorithm, based on calibration

The algorithm (a similar idea was used by Foster & Vohra [13]) that constructs an ε-internally consistent strategy is based on this simple fact: if the forecaster can, stage by stage, foresee the law of Nature's next action, say $y \in \Delta(\mathcal{J})$, then he just has to choose any best response to y at the following stage. The continuity of ρ implies that the forecasts need not be extremely precise but only up to some $\delta > 0$.

Let $\{y(l);\ l \in \mathcal{L}\}$ be a δ-grid of $\Delta(\mathcal{J})$ (*i.e.* a finite set such that for every $y \in \Delta(\mathcal{J})$ there exists $l \in \mathcal{L}$ such that $\|y - y(l)\| \leq \delta$) and $i(l)$ be a best response to $y(l)$, for every $l \in \mathcal{L}$. Then if δ is small enough:

$$\|y - y(l)\| \leq 2\delta \Rightarrow i(l) \in BR_{2\varepsilon}(y)$$

It is possible to construct a *good sequence of forecasts* by computing a calibrated strategy (introduced by Dawid [12] and recalled in the following subsection 1.2.1).

1.2.1 Calibration

Consider a two-person repeated game Γ_c where, at stage n, Nature chooses the state of the world j_n in a finite set $\mathcal{J}$ and a decision maker (that will be referred in this setting as a predictor) predicts it by choosing $y(l_n)$ in $\mathcal{Y} = \{y(l);\ l \in \mathcal{L}\}$, a finite δ-grid of $\Delta(\mathcal{J})$ – its cardinality is denoted by L. As usual, a behavioral strategy σ of the predictor (resp. τ of Nature) is a mapping from the set of finite histories $H = \bigcup_{n\in\mathbb{N}} (\mathcal{L} \times \mathcal{J})^n$ to $\Delta(\mathcal{L})$ (resp. $\Delta(\mathcal{J})$). We also denote by $\mathbb{P}_{\sigma,\tau}$ the probability generated by the pair (σ, τ) over $(\mathcal{H}, \mathcal{A})$ the set of infinite histories embedded with the cylinder topology.

Definition 1.4 (Dawid [12]) *A strategy σ of the predictor is calibrated (with respect to $\mathcal{Y} = \{y(l);\ l \in \mathcal{L}\}$) if for every strategy τ of Nature, $\mathbb{P}_{\sigma,\tau}$-as:*

$$\limsup_{n\to\infty} \frac{|N_n(l)|}{n}\left(\|\bar{j}_n(l) - y(l)\|^2 - \|\bar{j}_n(l) - y(k)\|^2 \right) \leq 0, \quad \forall k, l \in \mathcal{L},$$

where $\|\cdot\|$ is the Euclidian norm of $\mathbb{R}^J$.

In words, a strategy is calibrated if for every $l \in \mathcal{L}$, the empirical distribution of states, on the set of stages where $y(l)$ was predicted, is closer to $y(l)$ than to any other $y(k)$ (or the *frequency of* l, $|N_n(l)|/n$, is small).

Given a finite grid of $\Delta(\mathcal{J})$, the existence of calibrated strategies has been proved by Foster & Vohra [14] using either the Expected Brier Score or a minmax theorem (actually this second argument is acknowledged to Hart). We give here a construction, related but simpler than the one of Foster and Vohra, due to Sorin [30].

Proposition 1.5 (Foster & Vohra [14]) *For any finite grid $\mathcal{Y}$ of $\Delta(\mathcal{J})$, there exist calibrated strategies with respect to $\mathcal{Y}$ such that for every strategy τ of Nature:*

$$\mathbb{E}_{\sigma,\tau}\left[\max_{l,k\in\mathcal{L}} \frac{|N_n(l)|}{n}\left(\|\bar{j}_n(l) - y(l)\|^2 - \|\bar{j}_n(l) - y(k)\|^2 \right)\right] \leq O\left(\frac{1}{\sqrt{n}}\right).$$

Proof. Consider the auxiliary game where, at stage $n \in \mathbb{N}$, the predictor (resp. Nature) chooses $l_n \in \mathcal{L}$ (resp. $j_n \in \mathcal{J}$) and the vector payoff is the matrix $U_n \in \mathbb{R}^{L\times L}$ where

$$U_n^{lk} = \begin{cases} \|j_n - y(l)\|^2 - \|j_n - y(k)\|^2 & \text{if } l = l_n \\ 0 & \text{otherwise.} \end{cases}$$

A strategy σ is calibrated with respect to $\mathcal{L}$ if $\bar{U}_n$ converges to the negative orthant. Indeed for every $l, k \in \mathcal{L}$, the (l,k)-th coordinate of $\bar{U}_n$ is

$$\begin{aligned} \bar{U}_n^{lk} &= \frac{|N_n(l)|}{n} \frac{\sum_{m\in N_n(l)} \|j_m - y(l)\|^2 - \|j_m - y(k)\|^2}{|N_n(l)|} \\ &= \frac{|N_n(l)|}{n}\left(\|\bar{j}_n(l) - y(l)\|^2 - \|\bar{j}_n(l) - y(k)\|^2 \right). \end{aligned}$$

Denote by $\bar{U}_n^+ := \left\{\max\left(0, \bar{U}_n^{lk}\right)\right\}_{l,k\in\mathcal{L}} =: \bar{U}_n - \bar{U}_n^-$ the positive part of $\bar{U}_n$ and by $\lambda_n \in \Delta(\mathcal{L})$ any invariant measure of $\bar{U}_n^+$. We recall that λ is an invariant measure of a nonnegative matrix U if, for every $l \in \mathcal{L}$,

$$\sum_{k\in\mathcal{L}} \lambda(k) U^{kl} = \lambda(l) \sum_{k\in\mathcal{L}} U^{lk}.$$

Its existence is a consequence of Perron-Frobenius Theorem, see *e.g.* Seneta [28].

Define the strategy σ of the predictor inductively as follows. Choose arbitrarily $\sigma(\emptyset)$, the law of the first action and at stage $n+1$, play accordingly to any invariant measure of $\bar{U}_n^+$. We claim that this strategy is an approachability strategy of the negative orthant of $\mathbb{R}^{L\times L}$ because it satisfies Blackwell [5]'s sufficient condition:

$$\forall n \in \mathbb{N}, \langle \bar{U}_n - \bar{U}_n^-, \mathbb{E}_{\lambda_n}\left[U_{n+1}|j_{n+1}\right] - \bar{U}_n^- \rangle \leq 0.$$

Indeed, for every possible $j_{n+1} \in \mathcal{J}$:

$$\langle \bar{U}_n^+, \mathbb{E}_{\lambda_n}\left[U_{n+1}|j_{n+1}\right] \rangle = 0 = \langle \bar{U}_n^+, \bar{U}_n^- \rangle, \tag{3}$$

where the second equality follows from the definition of positive and negative parts.

Consider the first equality. The (l,k)-th coordinate of $\mathbb{E}_{\lambda_n}[U_{n+1}|j_{n+1}]$ is $\lambda_n(l)\left(\|j_{n+1}-y(l)\|^2 - \|j_{n+1}-y(k)\|^2\right)$, therefore the coefficient of $\|j_{n+1}-y(l)\|^2$ in the first term is $\lambda_n(l)\sum_{k\in\mathcal{L}}\left(\bar{U}_n^+\right)^{lk} - \sum_{k\in\mathcal{L}}\lambda_n(k)\left(\bar{U}_n^+\right)^{kl}$. This equals 0 since λ_n is an invariant measure of $\bar{U}_n^+$.

Blackwell [5]'s result also implies that $\mathbb{E}_{\sigma,\tau}\left[\|\bar{U}_n^+\|\right] \leq 2M_n n^{-1/2}$ for any strategy τ of Nature where $M_n^2 = \sup_{m\leq n}\mathbb{E}_{\sigma,\tau}\left[\|U_m\|^2\right] = 4L$. □

Interestingly, the strategy σ we constructed in this proof is actually internally consistent in the game with action spaces $\mathcal{L}$ and $\mathcal{J}$ and payoffs defined by $\rho(l,j) = -\|j-y(l)\|^2$.

Corollary 1.6 *For any finite grid $\mathcal{Y}$ of $\Delta(\mathcal{J})$, there exists σ, a calibrated strategy with respect to $\mathcal{Y}$, such that for every strategy τ of Nature, with $\mathbb{P}_{\sigma,\tau}$ probability at least $1-\delta$:*

$$\max_{l,k\in\mathcal{L}}\frac{|N_n(l)|}{n}\left(\|\bar{j}_n(l)-y(l)\|^2 - \|\bar{j}_n(l)-y(k)\|^2\right) \leq \frac{2M_n}{\sqrt{n}} + \Theta_n,$$

$$\text{where} \quad \Theta_n = \min\left\{\frac{v_n}{\sqrt{n}}\sqrt{2\ln\left(\frac{L^2}{\delta}\right)} + \frac{2}{3}\frac{K_n}{n}\ln\left(\frac{L^2}{\delta}\right), \frac{K_n}{\sqrt{n}}\sqrt{2\ln\left(\frac{L^2}{\delta}\right)}\right\};$$

$$M_n = \sup_{m\leq n}\sqrt{\mathbb{E}_{\sigma,\tau}\left[\|U_m\|^2\right]} \leq 3\sqrt{L};$$

$$v_n^2 = \sup_{m\leq n}\sup_{l,k\in\mathcal{L}}\mathbb{E}_{\sigma,\tau}\left[\left|U_n^{lk} - \mathbb{E}_{\sigma,\tau}\left[U_n^{lk}\right]\right|^2\right] \leq 3;$$

$$K_n = \sup_{m\leq n}\sup_{l,k\in\mathcal{L}}\left|U_n^{lk} - \mathbb{E}_{\sigma,\tau}\left[U_n^{lk}\right]\right| \leq 3.$$

Proof. Proposition 1.5 implies that $\mathbb{E}_{\sigma,\tau}\left[\bar{U}_n\right] \leq 2M_n n^{-1/2}$. Hoeffding-Azuma's inequality (see Lemma 3.4 below in section 3.3.1) implies that with probability at least $1-\delta$:

$$\bar{U}_n^{lk} - \mathbb{E}_{\sigma,\tau}\left[\bar{U}_n^{lk}\right] \leq \frac{K_n}{\sqrt{n}}\sqrt{2\ln\left(\frac{1}{\delta}\right)}.$$

Freedman's inequality (an analogue of Bernstein's inequality for martingale see Freedman [15], Proposition 2.1 or Cesa-Bianchi & Lugosi [9], Lemma A.8) implies that with probability at least $1-\delta$:

$$\bar{U}_n^{lk} - \mathbb{E}_{\sigma,\tau}\left[\bar{U}_n^{lk}\right] \leq \frac{v_n}{\sqrt{n}}\sqrt{2\ln\left(\frac{1}{\delta}\right)} + \frac{2}{3}\frac{K_n}{n}\ln\left(\frac{1}{\delta}\right).$$

The result is a consequence of these two inequalities and of Proposition 1.5. □

The definition of Θ_n as a minimum (and the use of Freedman's inequality) will be useful when we will refer to this corollary in the subsequent sections. Obviously, in the current framework, $\Theta_n \leq \frac{3}{\sqrt{n}}\sqrt{2\ln\left(\frac{L^2}{\delta}\right)}$.

1.2.2 Back to the Naïve Algorithm

Let us now go back to the construction of ε-consistent strategies in Γ. Compute σ, a calibrated strategy with respect to a δ-grid $\mathcal{Y} = \{y(l);\ l \in \mathcal{L}\}$ of $\Delta(\mathcal{J})$ in an abstract calibration game Γ_c. Whenever the decision maker (seen as a predictor) should choose the action l in Γ_c, then he (seen as a forecaster) chooses $i(l) \in BR(y(l))$ in the original game Γ. We claim that this defines a strategy σ_ε which is 2ε-internally consistent.

Proposition 1.7 (Foster & Vohra [13]) *For every $\varepsilon > 0$, the strategy σ_ε described above is 2ε-internally consistent.*

Proof. By definition of a calibrated strategy, for every $\eta > 0$, there exists with probability 1, an integer $N \in \mathbb{N}$ such that for every $l, k \in \mathcal{L}$ and for every $n \geq N$:

$$\frac{|N_n(l)|}{n}\Big(\|\bar{\jmath}_n(l) - y(l)\|^2 - \|\bar{\jmath}_n(l) - y(k)\|^2 \Big) \leq \eta.$$

Since $\{y(k);\ k \in \mathcal{L}\}$ is a δ-grid of $\Delta(\mathcal{J})$, for every $l \in \mathcal{L}$ and every $n \in \mathbb{N}$, there exists $k \in \mathcal{L}$ such that $\|\bar{\jmath}_n(l) - y(k)\|^2 \leq \delta^2$, hence $\|\bar{\jmath}_n(l) - y(l)\|^2 \leq \delta^2 + \eta\frac{n}{|N_n(l)|}$. Therefore, since $i(l) \in BR(y(l))$:

$$\frac{|N_n(l)|}{n} \geq \frac{\eta}{\delta^2} \Rightarrow \|\bar{\jmath}_n(l) - y(l)\|^2 \leq 2\delta^2 \Rightarrow \rho(k, \bar{\jmath}_n(l)) - \rho(i(l), \bar{\jmath}_n(l)) \leq 2\varepsilon,$$

for every $k \in \mathcal{I}$, $l \in \mathcal{L}$ and $n \geq N$. The (i,k)-th coordinate of $\bar{R}_n$ satisfies:

$$\begin{aligned}
\frac{|N_n(i)|}{n}\Big(\bar{R}_n^{ik} - 2\varepsilon\Big) &\leq \frac{1}{n}\sum_{m \in N_n(i)} \Big(\rho(k, j_m) - \rho(i, j_m) - 2\varepsilon\Big) \\
&= \frac{1}{n}\sum_{l:i(l)=i}\ \sum_{m \in N_n(l)} \Big(\rho(k, j_m) - \rho(i, j_m) - 2\varepsilon\Big) \\
&= \sum_{l:i(l)=i} \frac{|N_n(l)|}{n}\Big(\rho(k, \bar{\jmath}_n(l)) - \rho(i(l), \bar{\jmath}_n(l)) - 2\varepsilon\Big).
\end{aligned}$$

Recall that either $\frac{|N_n(l)|}{n} \geq \frac{\eta}{\delta^2}$ and $\rho(k, \bar{\jmath}_n(i)) - \rho(i(l), \bar{\jmath}_n(l)) - 2\varepsilon \leq 0$, or $\frac{|N_n(l)|}{n} < \frac{\eta}{\delta^2}$. Since ρ is bounded (by $M_\rho > 0$), then :

$$\frac{|N_n(i)|}{n}\Big(\bar{R}_n^{ik} - 2\varepsilon\Big) \leq \eta\frac{2M_\rho L}{\delta^2}, \quad \forall i \in \mathcal{I},\, \forall k \in \mathcal{I},\, \forall n \geq N,$$

which implies that σ is 2ε-internally consistent. □

Remark 1.8 *This naïve algorithm only achieves ε-consistency and Proposition 1.5 implies that*

$$\mathbb{E}_{\sigma,\tau}\left[\max_{i,k \in \mathcal{I}}\Big(\bar{R}_n^{ik} - \varepsilon\Big)\right] \leq O\left(\frac{1}{\sqrt{n}}\right).$$

The constants depend drastically on L, which is in the current framework in the order of ε^J, therefore it is not possible to obtain 0-internally consistency at the same rate with a classic doubling trick argument (i.e. use a

2^{-k}-internally consistent strategy on N_k stages, then switch to a $2^{-(k+1)}$-internally consistent strategy, and so on, see e.g. Sorin [29], Proposition 3.2 page 56).

Moreover, since this algorithm is based on calibration, it computes at each stage an invariant measure of a non-negative matrix; this can be done, using Gaussian elimination, with $O\left(L^3\right)$ operations, thus this algorithm is far from being efficient (since its computational complexity is polynomial in ε and exponential in J). There exist 0-internally consistent algorithms, see e.g. the reduction of Blum & Mansour [7], that do not have this exponential dependency in the complexity or in the constants.

On the bright side, this algorithm can be modified to obtain 0-consistency at optimal rate; obviously, it will still not be efficient with full monitoring (see section 1.4). However, it has to be understood as a tool that can be easily adapted in order to exhibit, in the partial monitoring case, an optimal internal consistent algorithm (see section 2.3). And in that last framework, it is not clear that we can remove the dependency on L (especially for the internal regret).

1.3 Calibration and Laguerre diagram

Given a finite subset of Voronoï sites $\{z(l) \in \mathbb{R}^d;\ l \in \mathcal{L}\}$, the l-th Voronoï cell $V(l)$, or the cell associated to $z(l)$, is the set of points closer to $z(l)$ than to any other $z(k)$:

$$V(l) = \left\{ Z \in \mathbb{R}^d;\ \|Z - z(l)\|^2 \leq \|Z - z(k)\|^2, \quad \forall k \in \mathcal{L} \right\},$$

where $\|\cdot\|$ is the Euclidian norm of $\mathbb{R}^d$. Each $V(l)$ is a polyhedron (as the intersection of a finite number of half-spaces) and $\{V(l);\ l \in \mathcal{L}\}$ is a covering of $\mathbb{R}^d$. A calibrated strategy with respect to $\{z(l);\ l \in \mathcal{L}\}$ has the property that for every $l \in \mathcal{L}$, the frequency of l goes to zero, or the empirical distribution of states on $N_n(l)$, converges to $V(l)$.

The naïve algorithm uses the Voronoï diagram associated to an arbitrary grid of $\Delta(\mathcal{J})$ and assigns to every small cell an ε-best reply to every point of it; this is possible by continuity of ρ. A calibrated strategy ensures that $\bar{j}_n(l)$ converges to $V(l)$ (or the frequency of l is small), thus choosing $i(l)$ on $N_n(l)$ was indeed a ε-best response to $\bar{j}_n(l)$. With this approach, we cannot construct immediately 0-internally consistent strategy. Indeed, this would require that for every $l \in \mathcal{L}$ there exists a 0-best response $i(l)$ to every element y in $V(l)$. However, there is no reason for them to share a common best response because $\{z(l);\ l \in \mathcal{L}\}$ is chosen arbitrarily.

On the other hand, consider the simple game called *Matching Penny*. Both players have two action *H*eads and *T*ails, so $\Delta(\mathcal{J}) = \Delta(\mathcal{I}) = [0,1]$, seen as the probability of choosing T. The payoff is 1 if both players choose the same action and -1 otherwise. Action H (resp. T) is a best response for Player 1 to any y in $[0,1/2]$ (resp. in $[1/2,1]$). These two segments are exactly the cells of the Voronoï diagram associated to $\{y(1) = 1/4, y(2) = 3/4\}$, therefore, performing a calibrated strategy with respect to $\{y(1), y(2)\}$ and playing H (resp. T) on the stages of type 1 (resp. 2) induces a 0-internally consistent strategy of Player 1.

This idea can be generalized to any game. Indeed, by Lemma 1.10 stated below, $\Delta(\mathcal{J})$ can be decomposed into polytopial best-response areas (a polytope is the convex hull of a finite number of points, its vertices). Given such a polytopial decomposition, one can find a finer Voronoï diagram (*i.e.* any best-response area is an union of Voronoï cells) and finally use a calibrated strategy to ensure convergence with respect to this diagram.

Although the construction of such a diagram is quite simple in $\mathbb{R}$, difficulties arise in higher dimension – even in $\mathbb{R}^2$. More importantly, the number of Voronoï sites can depend not only on the number of defining hyperplanes but also on the angles between them (thus being arbitrarily large even with a few hyperplanes). On the other hand, the description of a Laguerre diagram – this concept generalizes Voronoï diagrams – that refines a polytopial decomposition is quite simple and is described in Proposition 1.11 below. For this reason, we will consider from now on this kind of diagram (sometimes also called Power diagram) .

Given a subset of Laguerre sites $\{z(l) \in \mathbb{R}^d;\ l \in \mathcal{L}\}$ and weights $\{\omega(l) \in \mathbb{R};\ l \in \mathcal{L}\}$, the l-th Laguerre cell $P(l)$ is defined by:

$$P(l) = \left\{ Z \in \mathbb{R}^d;\ \|Z - z(l)\|^2 - \omega(l) \leq \|Z - z(k)\|^2 - \omega(k), \quad \forall k \in \mathcal{L} \right\},$$

where $\|\cdot\|$ is the Euclidian norm of $\mathbb{R}^d$. Each $P(l)$ is a polyhedron and $\mathcal{P} = \{P(l);\ l \in \mathcal{L}\}$ is a covering of $\mathbb{R}^d$.

Definition 1.9 *A covering $\mathcal{K} = \{K^i;\ i \in \mathcal{I}\}$ of a polytope K with non-empty interior is a polytopial complex of K if for every i, j in the finite set $\mathcal{I}$, K^i is a polytope with non-empty interior and the polytope $K^i \cap K^j$ has empty interior.*

This definition extends naturally to a polytope K with empty interior, if we consider the affine subspace generated by K.

Lemma 1.10 *There exists a subset $\mathcal{I}' \subset \mathcal{I}$ such that $\{B^i;\ i \in \mathcal{I}'\}$ is a polytopial complex of $\Delta(\mathcal{J})$, where B^i is the i-th best response area defined by*

$$B^i = \{y \in \Delta(\mathcal{J});\ i \in BR(y)\} = BR^{-1}(i).$$

Proof. For any $y \in \Delta(\mathcal{J})$, $\rho(\cdot, y)$ is linear on $\Delta(\mathcal{I})$ thus it attains its maximum on $\mathcal{I}$ and $\bigcup_{i\in\mathcal{I}} B^i = \Delta(\mathcal{J})$. Without loss of generality, we can assume that each B^i is non-empty, otherwise we drop the index i. For every $i, k \in \mathcal{I}$, $\rho(i,\cdot) - \rho(k,\cdot)$ is linear on $\Delta(\mathcal{J})$ therefore B^i is a polytope; it is indeed defined by

$$\begin{aligned} B^i &= \{y \in \Delta(\mathcal{J});\ \rho(i,y) \geq \rho(k,y),\ \forall k \in \mathcal{I}\} \\ &= \bigcap_{k\in\mathcal{I}} \{y \in \mathbb{R}^J;\ \rho(i,y) - \rho(k,y) \geq 0\} \cap \Delta(\mathcal{J}), \end{aligned}$$

so it is the intersection of a finite number of half-spaces and the polytope $\Delta(\mathcal{J})$.

Moreover if B_0^{ik}, the interior of $B^i \cap B^k$, is non-empty then $\rho(i,\cdot)$ equals $\rho(k,\cdot)$ on the subspace generated by B_0^{ik} and therefore on $\Delta(\mathcal{J})$; consequently $B^i = B^k$. Denote by $\mathcal{I}'$ any subset of $\mathcal{I}$ such that for every $i \in \mathcal{I}$, there exists exactly one $i' \in \mathcal{I}'$ such that $B^i = B^{i'} \neq \emptyset$, then $\{B^i;\ i \in \mathcal{I}'\}$ is a polytopial complex of $\Delta(\mathcal{J})$. □

Proposition 1.11 *Let $\mathcal{K} = \{K^i;\ i \in \mathcal{I}\}$ be a polytopial complex of a polytope $K \subset \mathbb{R}^d$. Then there exists $\{z(l) \in \mathbb{R}^d, \omega(l) \in \mathbb{R};\ l \in \mathcal{L}\}$, a finite set of Laguerre sites and weights, such that the Laguerre diagram $\mathcal{P} = \{P(l);\ l \in \mathcal{L}\}$ refines $\mathcal{K}$, i.e. every K^i is a finite union of cells.*

Proof. Let $\mathcal{K} = \{K^i;\ i \in \mathcal{I}\}$ be a polytopial complex of $K \subset \mathbb{R}^d$. Each K^i is a polytope, thus defined by a finite number of hyperplanes. Denote by $\mathcal{H} = \{H_t;\ t \in \mathcal{T}\}$ the set of all defining hyperplanes (the finite cardinality of $\mathcal{T}$ is denoted by T) and $\widehat{\mathcal{K}} = \{\widehat{K}^l;\ l \in \mathcal{L}\}$ the finest decomposition of $\mathbb{R}^d$ induced by $\mathcal{H}$ – usually called arrangement of hyperplanes – which by definition refines $\mathcal{K}$. Theorem 3 and Corollary 1 of Aurenhammer [2] imply that $\widehat{\mathcal{K}}$ is the Laguerre diagram associated to some $\{z(l), \omega(l);\ l \in \mathcal{L}\}$ whose exact computation requires the following notation:

i) for every $t \in \mathcal{T}$, let $c_t \in \mathbb{R}^d$ and $b_t \in \mathbb{R}$ (which can, without loss of generality, be assumed to be non zero) such that

$$H_t = \left\{X \in \mathbb{R}^d;\ \langle X, c_t\rangle = b_t\right\}.$$

ii) For every $l \in \mathcal{L}$ and $t \in \mathcal{T}$, $\sigma_t(l) = 1$ if the origin of $\mathbb{R}^d$ and $\widehat{K}^l$ are in the same halfspace defined by H_t and $\sigma_t(l) = -1$ otherwise.

iii) For every $l \in \mathcal{L}$, we define :

$$z(l) = \frac{\sum_{t\in\mathcal{T}} \sigma_t(l) c_t}{T} \quad \text{and} \quad \omega(l) = \|z(l)\|^2 + 2\frac{\sum_{t\in\mathcal{T}} \sigma_t(l) b_t}{T}. \tag{4}$$

Note that one can add the same constant to every weight $\omega(l)$. □

Buck [8] proved that the number of cells defined by T hyperplanes in $\mathbb{R}^d$ is bounded by $\sum_{k=0}^{d} \binom{T}{k} =: \phi(T, d)$, where $\binom{T}{k}$ is the binomial coefficient, T *choose* k. Moreover, T is smaller than $I(I-1)/2$ (in the case where each K^i has a non-empty intersection with every other polytope), so $L \leq \phi\left(\frac{I^2}{2}, d\right)$.

If $d \geq n$, then $\phi(n, d) = 2^n$. Pascal's rule and a simple induction imply that, for every $n, d \in \mathbb{N}$, $\phi(n, d) \leq (n+1)^d$. Finally, for any $n \geq 2d$, by noticing that

$$\frac{\binom{n}{d} + \binom{n}{d-1} + \ldots + \binom{n}{0}}{\binom{n}{d}} \leq \sum_{m=0}^{d} \left(\frac{d}{n-d+1}\right)^m \leq \sum_{m=0}^{\infty} \left(\frac{d}{n-d+1}\right)^m$$

which equals $\frac{n-d+1}{n-2d+1} \leq 1+d$, we deduce that $\phi(n, d) \leq (1+d)\binom{n}{d} \leq (1+d)\frac{n^d}{d!}$.

Lemma 1.12 *Let $\mathcal{P} = \{P(l);\ l \in \mathcal{L}\}$ be a Laguerre diagram associated to the set of sites and weights $\{z(l) \in \mathbb{R}^d, \omega(l) \in \mathbb{R};\ l \in \mathcal{L}\}$. Then, there exists a positive constant $M_P > 0$ such that for every $Z \in \mathbb{R}^d$ if*

$$\|Z - z(l)\|^2 - \omega(l) \leq \|Z - z(k)\|^2 - \omega(k) + \varepsilon, \quad \forall l, k \in \mathcal{L} \tag{5}$$

then $d(Z, P(l))$ is smaller than $M_P\varepsilon$.

The proof can be found in Appendix A.1; the constant M_P depends on the Laguerre diagram, and more precisely on the inner products $\langle c_t, c_{t'} \rangle$, for every $t, t' \in \mathcal{T}$.

1.4 Optimal algorithm with full monitoring

We reformulate Proposition 1.5 and Corollary 1.6 in terms of Laguerre diagram.

Theorem 1.13 *For any set of sites and weights* $\{y(l) \in \mathbb{R}^J,\, \omega(l) \in \mathbb{R};\ l \in \mathcal{L}\}$ *there exists a strategy* σ *of the predictor such that for every strategy* τ *of Nature:*

$$\mathbb{E}_{\sigma,\tau}\left[\left\|\left(\bar{U}_{\omega,n}\right)^+\right\|\right] \leq O\left(\frac{1}{\sqrt{n}}\right) \text{ where } U_{\omega,n} \text{ is defined by :}$$

$$U^{lk}_{\omega,n} = \begin{cases} \left[\|j_n - y(l)\|^2 - \omega(l)\right] - \left[\|j_n - y(k)\|^2 - \omega(k)\right] & \text{if } l = l_n \\ 0 & \text{otherwise} \end{cases}$$

Corollary 1.14 *For any set of sites and weights* $\{y(l) \in \mathbb{R}^J,\, \omega(l) \in \mathbb{R};\ l \in \mathcal{L}\}$*, there exists a strategy* σ *of the predictor such that, for every strategy* τ *of Nature, with* $\mathbb{P}_{\sigma,\tau}$ *probability at least* $1-\delta$*, and* $l, l \in \mathcal{L}$*:*

$$\frac{|N_n(l)|}{n}\left(\left[\|\bar{j}_n(l) - y(l)\|^2 - \omega(l)\right] - \left[\|\bar{j}_n(l) - y(k)\|^2 - \omega(k)\right]\right) \leq \frac{2M_n}{\sqrt{n}} + \Theta_n$$

$$\begin{aligned}
\text{where } M_n &= \sup_{m\leq n}\sqrt{\mathbb{E}_{\sigma,\tau}\left[\|U_{\omega,m}\|^2\right]} \leq 4\sqrt{L}\|(b,c)\|_\infty; \\
\Theta_n &= \min\left\{\frac{v_n}{\sqrt{n}}\sqrt{2\ln\left(\frac{L^2}{\delta}\right)} + \frac{2}{3}\frac{K_n}{n}\ln\left(\frac{L^2}{\delta}\right), \frac{K_n}{\sqrt{n}}\sqrt{2\ln\left(\frac{L^2}{\delta}\right)}\right\}; \\
v_n^2 &= \sup_{m\leq n}\sup_{l,k\in\mathcal{L}}\mathbb{E}_{\sigma,\tau}\left[\left|U^{lk}_{\omega,m} - \mathbb{E}_{\sigma,\tau}\left[U^{lk}_{\omega,m}\right]\right|^2\right] \leq 4\|(b,c)\|^2_\infty; \\
K_n &= \sup_{m\leq n}\sup_{l,k\in\mathcal{L}}\left|U^{lk}_{\omega,m} - \mathbb{E}_{\sigma,\tau}\left[U^{lk}_{\omega,m}\right]\right| \leq 4\|(b,c)\|_\infty, \\
\|(b,c)\|_\infty &= \sup_{t\in\mathcal{T}}\|c_t\| + \sup_{t\in\mathcal{T}}|b_t|.
\end{aligned}$$

Such a strategy is said to be calibrated with respect to $\{y(l),\, \omega(l);\ l \in \mathcal{L}\}$*.*

The proof are identical to the one of Proposition 1.5 and Corollary 1.6. We have now the material to construct our new *tool algorithm*:

Theorem 1.15 *There exists an internally consistent strategy* σ *of the forecaster such that for every strategy* τ *of Nature and every* $n \in \mathbb{N}$*, with* $\mathbb{P}_{\sigma,\tau}$ *probability greater than* $1-\delta$*:*

$$\max_{i,k\in\mathcal{I}}\bar{R}^{ik}_n \leq O\left(\sqrt{\frac{\ln\left(\frac{1}{\delta}\right)}{n}}\right). \tag{6}$$

Proof. The existence of a Laguerre Diagram $\{Y(l);\ l \in \mathcal{L}\}$ associated to a finite set $\{y(l) \in \mathbb{R}^J, \omega(l) \in \mathbb{R};\ l \in \mathcal{L}\}$ that refines $\{B^i;\ i \in \mathcal{I}\}$ is implied by Lemma 1.10 and Proposition 1.11. So, for every $l \in \mathcal{L}$, there exists $i(l)$ such that $Y(l) \subset B^{i(l)}$. As in the naïve algorithm, the strategy σ of the decision maker is constructed through a strategy $\widehat{\sigma}$ calibrated with respect to $\{y(l), \omega(l);\ l \in \mathcal{L}\}$. Whenever, accordingly to $\widehat{\sigma}$, the decision maker (seen as a predictor) should play l in Γ_c, then he (seen as a forecaster) plays $i(l)$ in Γ.

If we denote by $\widetilde{\jmath}_n(l)$ the projection of $\bar{\jmath}_n(l)$ onto $Y(l)$ then:

$$\begin{aligned}
\bar{R}_n^{ik} &= \sum_{l:i(l)=i} \frac{|N_n(l)|}{n}\Big(\rho\big(k,\bar{\jmath}_n(l)\big) - \rho\big(i(l),\bar{\jmath}_n(l)\big)\Big) \\
&\leq \sum_{l:i(l)=i} \frac{|N_n(l)|}{n}\Big(\Big[\rho\big(k,\bar{\jmath}_n(l)\big) - \rho\big(k,\widetilde{\jmath}_n(l)\big)\Big] \\
&\qquad + \Big[\rho\big(i(l),\widetilde{\jmath}_n(l)\big) - \rho\big(i(l),\bar{\jmath}_n(l)\big)\Big]\Big) \\
&\leq \sum_{l:i(l)=i} \frac{|N_n(l)|}{n}\Big(2M_\rho \left\|\widetilde{\jmath}_n(l) - \bar{\jmath}_n(l)\right\|\Big) \\
&\leq (2M_\rho M_P L) \max_{l,k\in\mathcal{L}} \frac{|N_n(l)|}{n}\Big(\Big[\|\bar{\jmath}_n(l) - y(l)\|^2 - \omega(l)\Big] \\
&\qquad - \Big[\|\bar{\jmath}_n(l) - y(k)\|^2 - \omega(k)\Big]\Big)
\end{aligned}$$

where the second inequality is due to the fact that $i(l) \in BR(\widetilde{\jmath}_n(l))$ and the third to the fact that ρ is M_ρ-Lipschitz. The fourth inequality is a consequence of Lemma 1.12.

Corollary 1.14 yields that for every strategy τ of Nature, with $\mathbb{P}_{\sigma,\tau}$ probability at least $1-\delta$:

$$\max_{l,k} \frac{N_n(l)}{n}\Big(\Big[\|\bar{\jmath}_n(l) - y(l)\|^2 - \omega(l)\Big] - \Big[\|\bar{\jmath}_n(l) - y(k)\|^2 - \omega(k)\Big]\Big) \leq \frac{8\sqrt{L}\|(b,c)\|_\infty}{\sqrt{n}} + \frac{4\|(b,c)\|_\infty}{\sqrt{n}}\sqrt{2\ln\left(\frac{L^2}{\delta}\right)},$$

therefore with $\Omega_0 = 16M_\rho M_P L^{3/2}\|(b,c)\|_\infty$ and $\Omega_1 = 8M_\rho M_P L^{1/2}\|(b,c)\|_\infty$ one has that for every strategy of Nature and with probability at least $1-\delta$:

$$\max_{i,k\in\mathcal{I}} \bar{R}_n^{ik} = \max_{i,k\in\mathcal{I}} \frac{|N_n(i)|}{n}\Big(\rho\big(k,\bar{\jmath}_n(i)\big) - \rho\big(i,\bar{\jmath}_n(i)\big)\Big) \leq \frac{\Omega_0}{\sqrt{n}} + \frac{\Omega_1}{\sqrt{n}}\sqrt{2\ln\left(\frac{L^2}{\delta}\right)}.$$

□

Remark 1.16 *Theorem 1.15 is already well-known. The construction of this internally consistent strategy relies on Theorem 1.13, which is implied by the existence of internally consistent strategies... Moreover, as mentioned before, it is far from being efficient since L – that enters both in the computational complexity and in the constant – is polynomial in I^J. There exist efficient algorithms, see e.g. Foster & Vohra [13] or Blum & Mansour [7].*

However, the calibration is defined in the space of Nature's action, where real payoffs are irrelevant; they are only used to decide which action is associated to each prediction. Therefore the algorithm does not require that the forecaster observes his real payoffs, as long as he knows what is the best response to his information (Nature's action in this case). This is precisely why our algorithm can be generalized to the partial monitoring framework.

The polytopial decomposition of $\Delta(\mathcal{J})$ induced by $\{b_t, c_t;\ t \in \mathcal{T}\}$ is exactly the same as the one induced by $\{\gamma b(t), \gamma c(t);\ t \in \mathcal{T}\}$ for any $\gamma > 0$. Thus, by choosing γ small enough, $\|(b,c)\|_\infty$ — and therefore the constants in Corollary 1.14 — can be arbitrarily small (*i.e.* multiplied by any $\gamma > 0$).

However, these two Laguerre diagrams are associated to the sets of sites and weights $\mathcal{L}(1)$ and $\mathcal{L}(\gamma)$, where $\mathcal{L}(\gamma) = \{\gamma z(l), \gamma\omega(l) + \gamma^2\|z(l)\|^2 - \gamma\|z(l)\|;\ l \in \mathcal{L}\}$. If $\mathcal{L}(\gamma)$ is used instead of $\mathcal{L}(1)$, then the constant M_P defined in Lemma 1.12 should be divided by γ. So, as expected, the constants in the proof of Theorem 1.15 do not depend on γ. From now on, we will assume that $\|(b,c)\|_\infty$ is smaller than 1.

2 Partial monitoring

2.1 Definitions

In the partial monitoring framework, the decision maker does not observe Nature's actions. There is a finite set of signals $\mathcal{S}$ (of cardinality S) such that, at stage n the forecaster receives only a random signal $s_n \in \mathcal{S}$. Its law is $s(i_n, j_n)$ where s is a mapping from $\mathcal{I} \times \mathcal{J}$ to $\Delta(\mathcal{S})$, known by the decision maker.

We define $\mathbf{s}$ from $\Delta(\mathcal{J})$ to $\Delta(\mathcal{S})^I$ by $\mathbf{s}(y) = \Big(\mathbb{E}_y\left[s(i,j)\right]\Big)_{i\in\mathcal{I}} \in \Delta(\mathcal{S})^I$. Any element of $\Delta(\mathcal{S})^I$ is called a flag (it is a vector of probability distributions over $\mathcal{S}$) and we will denote by $\mathcal{F}$ the range of $\mathbf{s}$. Given a flag f in $\mathcal{F}$, the decision maker cannot distinguish between any different mixed actions y and y' in $\Delta(\mathcal{J})$ that generate f, *i.e.* such that $\mathbf{s}(y) = \mathbf{s}(y') = f$. Thus $\mathbf{s}$

is the maximal informative mapping about Nature's action. We denote by $f_n = \mathbf{s}(j_n)$ the (unobserved) flag of stage $n \in \mathbb{N}$.

Example 2.1 *Label efficient prediction (Example 6.8 in Cesa-Bianchi & Lugosi [9]):*

Consider the following game. Nature chooses an outcome G or B and the forecaster can either observe the actual outcome (action o) or choose to not observe it and pick a label g or b. His payoff is equal to 1 if he chooses the right label and otherwise is equal to 0. Payoffs and laws of signals are defined by the following matrices (where a, b and c are three different probabilities over a finite given set S).

Payoffs:

	G	B
o	0	0
g	0	1
b	1	0

and signals:

	G	B
o	a	b
g	c	c
b	c	c

Action G, whose best response is g, generates the flag (a, c, c) and action B, whose best response is b, generates the flag (b, c, c). In order to distinguish between those two actions, the forecaster needs to know $s(o, y)$ although action o is never a best response (but is purely informative).

The worst payoff compatible with x and $f \in \mathcal{F}$ is defined by:

$$W(x, f) = \inf_{y \in \mathbf{s}^{-1}(f)} \rho(x, y), \tag{7}$$

and W is extended to $\Delta(\mathcal{S})^I$ by $W(x, f) = W\left(x, \Pi_{\mathcal{F}}(f)\right)$.

As in the full monitoring case, we define, for every $\varepsilon \geq 0$, the ε-best response multivalued mapping $BR_\varepsilon : \Delta(\mathcal{S})^I \rightrightarrows \Delta(\mathcal{I})$ by :

$$BR_\varepsilon(f) = \left\{ x \in \Delta(\mathcal{I});\ W(x, f) \geq \sup_{z \in \Delta(\mathcal{I})} W(z, f) - \varepsilon \right\}.$$

Given a flag $f \in \Delta(\mathcal{S})^I$, the function $W(\cdot, f)$ may not be linear so the best response of the forecaster might not contain any element of $\mathcal{I}$.

Example 2.2 *Matching Penny in the dark:*

Consider the Matching Penny game where the forecaster does not observe the coin but always receives the same signal c: every choice of Nature generates the same flag (c, c). For every $x \in [0, 1] = \Delta(\{H, T\})$ – the probability of playing T –, the worst compatible payoff $W(x, (c, c)) = \min_{y \in \Delta(J)} \rho(x, y)$

is equal to $-|1-2x|$ *thus is non-negative only for* $x = 1/2$*. Therefore the only best response of the forecaster is to play* $\frac{1}{2}H + \frac{1}{2}T$*, while actions* H *and* T *give the worst payoff of -1.*

The definition of external consistency and especially equation (1) extend naturally to this framework: a strategy of the decision maker is externally consistent if he could not have improved his payoff by knowing, before the beginning of the game, the average flag:

Definition 2.3 (Rustichini [27]) *A strategy* σ *of the forecaster is externally consistent if for every strategy* τ *of Nature:*

$$\limsup_{n\to+\infty} \max_{z\in\Delta(\mathcal{I})} W(z, \bar{f}_n) - \bar{\rho}_n \leq 0, \quad \mathbb{P}_{\sigma,\tau}\text{-}as.$$

The main issue is the definition of internally consistency. In the full monitoring case, the forecaster has no internal regret if, for every $i \in \mathcal{I}$, the action i is a best-response to the empirical distribution of Nature's actions, on the set of stages where i was actually chosen. In the partial monitoring framework, the decision maker's action should be a best response to the average flag. Since it might not belong to $\mathcal{I}$ but rather to $\Delta(\mathcal{I})$, we will (following Lehrer & Solan [21]) distinguish the stages not as a function of the action actually chosen, but as a function of its law.

We make an extra assumption on the characterization of the forecaster's strategy: it can be generated by a finite family of mixed actions $\{x(l) \in \Delta(\mathcal{I});\ l \in \mathcal{L}\}$ such that, at stage $n \in \mathbb{N}$, the forecaster chooses a type l_n and, given that type, the law of his action i_n is $x(l_n) \in \Delta(\mathcal{I})$.

Denote by $N_n(l) = \{m \in \{1,\dots,n\};\ l_m = l\}$ the set of stages before the n-th whose type is l. Roughly speaking, a strategy will be ε-internally consistent (with respect to the set $\mathcal{L}$) if, for every $l \in \mathcal{L}$, $x(l)$ is an ε-best response to $\bar{f}_n(l)$, the average flag on $N_n(l)$ (or the frequency of the type l, $|N_n(l)|/n$, converges to zero).

The finiteness of $\mathcal{L}$ is required to get rid of strategies that trivially insure that every frequency converges to zero (for instance by choosing only once every mixed action). The choice of $\{x(l);\ l \in \mathcal{L}\}$ and the description of the strategies are justified more precisely below by Remark 2.7 in section 2.3.

Definition 2.4 (Lehrer & Solan [21]) *For every* $n \in \mathbb{N}$ *and every* $l \in \mathcal{L}$*, the average internal regret of type* l *at stage* n *is*

$$\mathcal{R}_n(l) = \sup_{x\in\Delta(\mathcal{I})} \left[W(x, \bar{f}_n(l)) - \bar{\rho}_n(l)\right].$$

A strategy σ of the forecaster is $(\mathcal{L},\varepsilon)$-internally consistent if for every strategy τ of Nature:

$$\limsup_{n\to+\infty}\frac{|N_n(l)|}{n}\Big(\mathcal{R}_n(l)-\varepsilon\Big)\leq 0,\quad \forall l\in\mathcal{L},\quad \mathbb{P}_{\sigma,\tau}\text{-}as.$$

In words, a strategy is $(\mathcal{L},\varepsilon)$-internally consistent if, for every $l\in\mathcal{L}$, the forecaster could not have had, for sure, a better payoff (of at least ε) if he had known, before the beginning of the game, the average flag on $N_n(l)$ (or the frequency of l is small).

2.2 A naïve algorithm

Theorem 2.5 (Lehrer & Solan [21]) *For every $\varepsilon>0$, there exist $(\mathcal{L},\varepsilon)$-internally consistent strategies.*

Lehrer & Solan [21] proved the existence and constructed such strategies and an alternative, yet close, algorithm has been provided by Perchet [24]. The main ideas behind them are similar to the full monitoring case so we will quickly describe them. For simplicity, we assume in the following sketch of the proof, that the decision maker fully observes the sequence of flags $f_n=\mathbf{s}(j_n)\in\Delta(\mathcal{S})^I$.

Recall that W is continuous (see Lugosi, Mannor & Stoltz [23], Proposition A.1), so for every $\varepsilon>0$ there exist two finite families $\mathcal{G}=\{f(l)\in\Delta(\mathcal{S})^I;\ l\in\mathcal{L}\}$, a δ-grid of $\Delta(\mathcal{S})^I$, and $X=\{x(l)\in\Delta(I);\ l\in\mathcal{L}\}$ such that if f is δ-close to $f(l)$ and x is δ-close to $x(l)$ then x belongs to $BR_\varepsilon(f)$. A calibrated algorithm ensures that:

i) $\bar{f}_n(l)$ is asymptotically δ-close to $f(l)$ - because it is closer to $f(l)$ than to every other $f(k)$;

ii) $\bar{\imath}_n(l)$ converges to $x(l)$ as soon as $|N_n(l)|$ is big enough - because on $N_n(l)$ the choices of action of the decision maker are independent and identically distributed accordingly to $x(l)$;

iii) $\bar{\rho}_n(l)$ converges to $\rho(x(l),\bar{\jmath}_n(l))$ which is greater than $W\Big(x(l),\bar{f}_n(l)\Big)$ — because $\bar{\jmath}_n(l)$ generates the flag $\bar{f}_n(l)$.

Therefore, $W\Big(x(l),\bar{f}_n(l)\Big)$ is close to $W\Big(x(l),f(l)\Big)$ which is greater than $W\Big(z,f(l)\Big)$ for any $z\in\Delta(\mathcal{I})$. As a consequence $\bar{\rho}_n(l)$ is asymptotically

greater (up to some $\varepsilon > 0$) than $\sup_z W\Big(z, \bar{f}_n(l)\Big)$, as long as $|N_n(l)|$ is big enough.

The difference between the two algorithm lies in the construction of a calibrated strategy. On one hand, the algorithm of Lehrer & Solan [21] reduces to Blackwell's approachability of some convex set $\mathcal{C} \subset \mathbb{R}^{LSI}$; it therefore requires to solve at each stage a linear program of size polynomial in ε^{SI}, after a projection on $\mathcal{C}$. On the other hand, the algorithm of Perchet [24] is based on the construction given in section 1.2.1; it solves at each stage a system of linear equation of size also polynomial in ε^{SI}.

The conclusions of the full monitoring case also apply here: these highly non-efficient algorithms cannot be used directly to construct $(\mathcal{L}, 0)$-internally consistent strategy with optimal rates since the constants depend drastically on ε . We will rather prove that one can define wisely once for all $\{f(l), \omega(l);\ l \in \mathcal{L}\}$ and $\{x(l);\ l \in \mathcal{L}\}$ (see Proposition 2.6 and Proposition 1.11) so that $x(l) \in \Delta(\mathcal{I})$ is a 0-best response to any flag f in $P(l)$, the Laguerre cell associated to $f(l)$ and $\omega(l)$.

The strategy associated with these choices will be $(\mathcal{L}, 0)$-internally consistent, with an optimal rate of convergence and a computational complexity polynomial in L.

2.3 Optimal algorithms

As in the full monitoring framework (cf Lemma 1.10), we define for every $x \in \Delta(\mathcal{I})$ the x-best response area B^x as the set of flags to which x is a best response :

$$B^x = \big\{f \in \Delta(\mathcal{S})^I;\ x \in BR(f)\big\} = BR^{-1}(x).$$

Since W is continuous, the family $\{B^x;\ x \in \Delta(\mathcal{I})\}$ is a covering of $\Delta(\mathcal{S})^I$. However, one of its finite subsets can be decomposed into a finite polytopial complex:

Proposition 2.6 *There exists a finite family $X = \{x(l) \in \Delta(\mathcal{I});\ l \in \mathcal{L}\}$ such that the family $\big\{B^{x(l)};\ l \in \mathcal{L}\big\}$ of associated best response area can be further subdivided into a polytopial complex of $\Delta(\mathcal{S})^I$.*

The rather technical proof can be found in Appendix A.2. In this framework and because of the lack of linearity of W, any $B^{x(l)}$ might not be convex nor connected. However, each one of them is a finite union of polytopes and the family of all those polytopes is a complex of $\Delta(\mathcal{S})^I$.

Remark 2.7 *As a consequence of Proposition 2.6, there exists a finite set $X \subset \Delta(\mathcal{I})$ that contains a best response to any flag f. In particular, if the decision maker could observe the flag f_n before choosing his action x_n then, at every stage, x_n would be in X. So in the description of the strategies of the forecaster, the finite set $\{x(l);\ l \in \mathcal{L}\} = X$ is in fact intrinsic i.e. determined by the description of the payoff and signal functions.*

As a consequence of this remark, mentioning $\mathcal{L}$ is irrelevant; so we will, from now on, simply speak of *internally consistent strategies*.

2.3.1 Outcome dependent signals

In this section, we assume that the laws of the signal received by the decision maker are independent of his action. Formally, for every $i, i' \in \mathcal{I}$, the two mappings $s(i,\cdot)$ and $s(i',\cdot)$ are equal. Therefore, $\mathcal{F}$ (the set of realizable flags) can be seen as a polytopial subset of $\Delta(\mathcal{S})$. Proposition 2.6 holds in this framework, hence there exists a finite family $\{x(l);\ l \in \mathcal{L}\}$ such that for any flag $f \in \mathcal{F}$, there is some $l \in \mathcal{L}$ such that $x(l)$ is a best-reply to f. Moreover, for a fixed $l \in \mathcal{L}$, the set of such flags is a polytope.

Theorem 2.8 *There exists an internally consistent strategy σ such that for every strategy τ of Nature, with $\mathbb{P}_{\sigma,\tau}$-probability at least $1-\delta$:*

$$\sup_{l\in\mathcal{L}} \frac{|N_n(l)|}{n}\mathcal{R}_n(l) \leq O\left(\sqrt{\frac{\ln\left(\frac{1}{\delta}\right)}{n}}\right). \tag{8}$$

Proof. Propositions 1.11 and 2.6 imply the existence of two finite families $\{x(l);\ l \in \mathcal{L}\}$ and $\{f(l), \omega(l);\ l \in \mathcal{L}\}$ such that $x(l)$ is a best response to any f in $P(l)$, the Laguerre cell associated to $f(l)$ and $\omega(l)$. Assume, for the moment, that for any two different l and k in $\mathcal{L}$, the probability measures $x(l)$ and $x(k)$ are different.

The strategy σ is defined as follows. Compute a strategy $\widehat{\sigma}$ calibrated with respect to $\{f(l), \omega(l);\ l \in \mathcal{L}\}$. When the decision maker (seen as a predictor) should choose $l \in \mathcal{L}$ accordingly to $\widehat{\sigma}$, then he (seen as a forecaster) plays accordingly to $x(l)$ in the original game. Corollary 1.14 (with the assumption that $\|(b,c)\|_\infty$ is smaller than 1) implies that with $\mathbb{P}_{\sigma,\tau}$ probability

at least $1-\delta_1$:

$$\max_{l\in\mathcal{L}}\frac{|N_n(l)|}{n}\Big(\Big[\|\bar{s}_n(l)-f(l)\|^2-\omega(l)\Big]-\Big[\|\bar{s}_n(l)-f(k)\|^2-\omega(k)\Big]\Big)\leq \frac{8\sqrt{L}}{\sqrt{n}}+\frac{4}{\sqrt{n}}\sqrt{2\ln\left(\frac{L^2}{\delta_1}\right)},$$

therefore combined with Lemma 1.12, this yields that :

$$\max_{l\in\mathcal{L}}\frac{|N_n(l)|}{n}\left\|\bar{s}_n(l)-\widetilde{f}_n(l)\right\|\leq\frac{8M_P\sqrt{L}}{\sqrt{n}}+\frac{4M_P}{\sqrt{n}}\sqrt{2\ln\left(\frac{L^2}{\delta_1}\right)},\tag{9}$$

where $\widetilde{f}_n(l)$ is the projection of $\bar{s}_n(l)$ onto $P(l)$.

Hoeffding-Azuma's inequality implies that with $\mathbb{P}_{\sigma,\tau}$ probability at least $1-\delta_2$:

$$\max_{l\in\mathcal{L}}\frac{|N_n(l)|}{n}\left\|\bar{s}_n(l)-\bar{f}_n(l)\right\|\leq\sqrt{\frac{2\ln\left(\frac{2SL}{\delta_2}\right)}{n}}\tag{10}$$

and with probability at least $1-\delta_3$:

$$\max_{l\in\mathcal{L}}\frac{|N_n(l)|}{n}\left|\bar{\rho}_n(l)-\rho(x(l),\bar{\jmath}_n(l))\right|\leq M_\rho\sqrt{\frac{2\ln\left(\frac{2L}{\delta_3}\right)}{n}}.\tag{11}$$

W is M_W-Lipschitz in f (see Lugosi, Mannor & Stoltz [23]) and $\mathbf{s}\left(\bar{\jmath}_n(l)\right)=\bar{f}_n(l)$ therefore:

$$\bar{\rho}_n(l)\geq W\Big(x(l),\widetilde{f}_n(l)\Big)-\left|\bar{\rho}_n(l)-\rho(x(l),\bar{\jmath}_n(l))\right|-M_W\left\|\bar{f}_n(l)-\widetilde{f}_n(l)\right\|\tag{12}$$

and $\max_{x\in\Delta(\mathcal{I})}W\left(x,\bar{f}_n(l)\right)$ is smaller than

$$\begin{aligned}&\max_{x\in\Delta(\mathcal{I})}W\Big(x,\widetilde{f}_n(l)\Big)+M_W\left(\left\|\bar{s}_n(l)-\bar{f}_n(l)\right\|+\left\|\bar{s}_n(l)-\widetilde{f}_n(l)\right\|\right)\\&=W\Big(x(l),\widetilde{f}_n(l)\Big)+M_W\left(\left\|\bar{s}_n(l)-\bar{f}_n(l)\right\|+\left\|\bar{s}_n(l)-\widetilde{f}_n(l)\right\|\right)\end{aligned}\tag{13}$$

since $x(l)$ is a best response to $\widetilde{f}_n(l)$. Equations (12) and (13) yield

$$\mathcal{R}_n(l)\leq 2M_W\left\|\bar{s}_n(l)-\bar{f}_n(l)\right\|+2M_W\left\|\bar{s}_n(l)-\widetilde{f}_n(l)\right\|+\left|\bar{\rho}_n(l)-\rho(x(l),\bar{\jmath}_n(l))\right|.\tag{14}$$

Combining equations (9), (10), (11) and (14) gives that with probability at least $1-\delta$, if we define $\Omega_0 = 16 M_P M_W \sqrt{L}$, $\Omega_1 = (2M_W + 8M_W M_P + M_\rho)$ and $\Omega_2 = L\,(L+2S+2)$:

$$\sup_{l\in\mathcal{L}} \frac{|N_n(l)|}{n}\mathcal{R}_n(l) \leq \frac{\Omega_0}{\sqrt{n}} + \frac{\Omega_1}{\sqrt{n}}\sqrt{2\ln\left(\frac{2\Omega_2}{\delta}\right)} \tag{15}$$

If there exist l and k such that $x(l) = x(k)$, then although the decision maker made two different predictions $f(l)$ or $f(k)$, he played accordingly to the same probability $x(l) = x(k)$. Define $N_n(l,k)$ as the set of stages where the decision maker predicts either $f(l)$ or $f(k)$ up to stage n, $\bar{f}_n(l,k)$ as the average flag on this set, $\bar{\rho}_n(l,k)$ as the average payoff and $\mathcal{R}_n(l,k)$ as the regret. Since $W(x,\cdot)$ is convex for every $x \in \Delta(\mathcal{I})$, then $\max_{x\in\Delta(\mathcal{I})} W(x,\cdot)$ is also convex so $\frac{|N_n(l,k)|}{n}\max_{x\in\Delta(\mathcal{I})} W(x,\bar{f}_n(l,k))$ is smaller than

$$\frac{|N_n(l)|}{n}\max_{x\in\Delta(\mathcal{I})} W(x,\bar{f}_n(l)) + \frac{|N_n(k)|}{n}\max_{x\in\Delta(\mathcal{I})} W(x,\bar{f}_n(k))$$

$$\text{and} \quad -\frac{|N_n(l,k)|}{n}\bar{\rho}_n(l,k) = -\frac{|N_n(l)|}{n}\bar{\rho}_n(l) - \frac{|N_n(k)|}{n}\bar{\rho}_n(k)$$

so we still have

$$\frac{|N_n(l,k)|}{n}\mathcal{R}_n(l,k) \leq O\left(\sqrt{\frac{\ln\left(\frac{1}{\delta}\right)}{n}}\right).$$

Hence the previous bound holds up to a factor L. □

Remark 2.9 *Lugosi, Mannor & Stoltz [23] have constructed an externally consistent strategy, i.e. such that, asymptotically, for any strategy τ of Nature:*

$$\bar{\rho}_n \geq \max_{z\in\Delta(\mathcal{I})} W\left(z,\bar{f}_n\right), \quad \mathbb{P}_{\sigma,\tau}-\text{as.}$$

The final argument in the proof of Theorem 2.8 also implies that an internally consistent strategy is also externally consistent, hence we can compare bounds between our algorithm.

If the signals are deterministic, Lugosi, Mannor & Stoltz [23]'s efficient algorithm has an expected regret smaller than $O\left(n^{-1/2}\right)$. However this bound became, with random signals, $O\left(n^{-1/4}\right)$. Thus our algorithm, along with computing no internal regret, has a better rate of convergence – the optimal one. Concerning the computational complexity, the true purpose of this algorithm being the minimization of internal regret, it is not efficient to bound external regret.

2.3.2 Action-Outcome dependant signals

In this section, we consider the most general framework and we assume that the laws of the signals might depend on the decision maker's actions. Our main result is the following:

Theorem 2.10 *There exists an internally consistent strategy σ such that, for every strategy τ of Nature, with $\mathbb{P}_{\sigma,\tau}$ probability at least $1-\delta$:*

$$\max_{l\in\mathcal{L}} \frac{|N_n(l)|}{n}\mathcal{R}_n(l) \leq O\left(\frac{1}{n^{1/3}}\sqrt{\ln\left(\frac{1}{\delta}\right)} + \frac{1}{n^{2/3}}\ln\left(\frac{1}{\delta}\right)\right). \tag{16}$$

Proof. The proof is essentially the same as the one of Theorem 2.8, so we can assume that $x(l) \neq x(k)$ for any two different l and k in $\mathcal{L}$. The only difference is due to the fact that at stage $n \in \mathbb{N}$, the unobserved flag f_n has to be estimated (see *e.g.* Lugosi, Mannor & Stoltz [23]).

Following Auer, Cesa-Bianchi, Freund & Schapire [1], we define for every $l \in \mathcal{L}$ and $n \in \mathbb{N}$, the $\widehat{\gamma}_n$-perturbation of $x(l)$ by $\widehat{x}(l,n) = (1-\widehat{\gamma}_n)x(l) + \widehat{\gamma}_n u$ where u is the uniform probability over $\mathcal{I}$ and $(\widehat{\gamma}_n)_{n\in\mathbb{N}}$ is a non-negative non-increasing sequence. For every $n \in \mathbb{N}$, let

$$e_n = \left(\frac{\mathbb{1}_{i=i_n}}{\widehat{x}(l_n,n)[i_n]}\left(\mathbb{1}_{s=s_n}\right)_{s\in\mathcal{S}}\right)_{i\in\mathcal{I}} \in \left(\mathbb{R}^S\right)^I,$$

where $\widehat{x}(l_n,n)[i_n] \geq \gamma_n = \widehat{\gamma}_n/I > 0$ is the weight put by $\widehat{x}(l_n,n)$ on i_n. With this notation, e_n is an unbiased estimator of f_n since $\mathbb{E}_{\sigma,\tau}\left[e_n|h^{n-1}\right] = f_n$, seen as an element of $\left(\mathbb{R}^S\right)^I$.

We define now the strategy of the forecaster. Assume that in an auxiliary game Γ_c, a predictor computes $\widetilde{\sigma}$, a calibrated strategy with respect to $\{f(l),\, \omega(l);\, l \in \mathcal{L}\}$, but where the state at stage n is the estimator $e_n \in \mathbb{R}^{IS}$. When the decision maker (seen as a predictor) should choose l_n accordingly to $\widetilde{\sigma}$ in Γ_c, then he (seen as a forecaster) chooses i_n accordingly to $\widehat{x}(l_n)$ in the original game.

In order to use Corollary 1.14, we need to bound v_n, M_n and K_n. In the current framework and thanks to Proposition 1.11, one has for every $l,k \in \mathcal{L}$ and $n \in \mathbb{N}$:

$$U^{l,k}_{\omega,n} = 2\mathbb{1}_{l=l_n}\sum_{t\in\mathcal{T}}\frac{\sigma_t(k)-\sigma_t(l)}{T}\Big(\langle e_n, c_t\rangle + b_t\Big),$$

so using the fact that $\|(b,c)\|_\infty^2 = 1$ and the definition of e_n:

$$\sup_{l,k\in\mathcal{L}}\sup_{m\leq n}\mathbb{E}_{\sigma,\tau}\left[\left|U_{\omega,m}^{l,k}\right|^2\right] \leq 16\mathbb{E}_{\sigma,\tau}\left[\|e_n\|^2\right] \leq 16\sum_{i\in\mathcal{I}}\frac{\widehat{x}(l_n,n)[i]}{(\widehat{x}(l_n,n)[i])^2} \leq 16\frac{I}{\gamma_n}.$$

As a consequence, $K_n \leq 4\frac{1}{\gamma_n}$, $v_n \leq 4\sqrt{\frac{I}{\gamma_n}}$ and $M_n \leq 4\sqrt{\frac{LI}{\gamma_n}}$. Lemma 1.12 implies that, with $\mathbb{P}_{\sigma,\tau}$ probability at least $(1-\delta_1)$, for every $l \in \mathcal{L}$:

$$\frac{|N_n(l)|}{n}\left\|\bar{e}_n(l) - \widetilde{f}_n(l)\right\| \leq \frac{8\sqrt{LI}M_P}{\sqrt{\gamma_n n}} + \frac{8\sqrt{I}M_P}{\sqrt{\gamma_n n}}\sqrt{2\ln\left(\frac{L^2}{\delta_1}\right)} + \frac{8}{3}\frac{M_P}{\gamma_n n}\ln\left(\frac{L^2}{\delta_1}\right),$$

where $\widetilde{f}_n(l)$ is the projection of $\bar{e}_n(l)$ onto $P(l)$.

Following Lugosi, Mannor & Stoltz [23], since for every $i \in \mathcal{I}$ and $s \in \mathcal{S}$, $\mathbb{E}_{\sigma,\tau}\left[|e_n^{i,s}|^2\right] \leq 1/\gamma_n$, Freedman's inequality implies that with probability at least $1-\delta_2$, for every $l \in \mathcal{L}$

$$\frac{|N_n(l)|}{n}\left\|\bar{e}_n(l) - \bar{f}_n(l)\right\| \leq \sqrt{IS}\left(\sqrt{2\frac{1}{n\gamma_n}\ln\left(\frac{2LIS}{\delta_2}\right)} + \frac{2}{3n\gamma_n}\ln\left(\frac{2LIS}{\delta_2}\right)\right).$$

Hoeffding-Azuma's inequality implies that with probability at least $1-\delta_3$:

$$\max_{l\in\mathcal{L}}\frac{N_n(l)}{n}\left|\bar{\rho}_n(l) - \rho(x(l),\bar{\jmath}_n(l))\right| \leq M_\rho\sqrt{\frac{2}{n}\ln\left(\frac{2L}{\delta_3}\right)} + 2M_\rho\frac{\sum_{m\in N_n(l)}\widehat{\gamma}_m}{n},$$

and by taking $\gamma_n = n^{-1/3}$, one has $\sum_{m\in N_n(l)}\widehat{\gamma}_m \leq \frac{3I}{2}n^{2/3}$. As a consequence, for every $l \in \mathcal{L}$, with probability at least $1-\delta$:

$$\frac{N_n(l)}{n}\mathcal{R}_n(l) \leq \frac{\Omega_1}{n^{1/3}} + \frac{\Omega_2}{n^{1/3}}\sqrt{2\ln\left(\frac{2\Omega_5}{\delta}\right)} + \frac{\Omega_3}{n^{1/2}}\sqrt{2\ln\left(\frac{2\Omega_5}{\delta}\right)} + \frac{2}{3}\frac{\Omega_4}{n^{2/3}}\ln\left(\frac{2\Omega_5}{\delta}\right)$$

with the constants defined by $\Omega_1 = 16M_PM_W\sqrt{LI} + 3M_WM_\rho I$, $\Omega_2 = 2M_W\sqrt{I}\left(8M_P + \sqrt{S}\right)$, $\Omega_3 = M_\rho$, $\Omega_4 = 2M_W(4M_P + \sqrt{IS})$ and $\Omega_5 = L\,(L + 2 + 2IS)$. They can be decreased if concentration inequalities in Hilbert spaces are used (see section 3.3). □

In the label efficient prediction game defined in Example 2.1, for every strategy σ of the decision maker there exists a sequence of outcomes such that the forecaster expected regret is greater than $n^{-1/3}/7$ (see Theorem 5.1 in Cesa-Bianchi, Lugosi & Stoltz [10]). Therefore the rate of $n^{-1/3}$ of our algorithm is optimal for both internal and external regret.

The computational complexity of this internally consistent algorithm is polynomial in L. Thus it can be seen, in some sense, as an efficient one. A question left open is the existence of an algorithm whose computational complexity is polynomial in the minimal number of best-response areas required to cover $\Delta(\mathcal{S})^I$, see Proposition 2.6.

The following section 3.1 deals with a simpler question and exhibits an internally consistent algorithm which requires to solve at each stage a linear program of size polynomial in L_0, the minimal number of polytopes on which BR is constant, instead of a system of linear equations of size L.

3 Concluding remarks

3.1 Second algorithm: calibration and polytopial complex.

The algorithms we described are quite easy to run stage by stage since the forecaster only needs to compute some invariant measures of non-negative matrices. However, they require to construct the Laguerre diagram $\mathcal{P} = \{P(l);\ l \in \mathcal{L}\}$ given the set $\{b_t,\, c_t;\ t \in \mathcal{T}\}$. And we have shown that L, which is a factor both in the complexity of the algorithms and in their rate of convergence, can be in the order of T^{SI} hence polynomial in L_0^{SI}.

This section is devoted to a modification of the algorithm that does not require to compute a Laguerre diagram but which is more difficult, stage by stage, to implement. The only difference between the two algorithms is in the definition of calibration.

Let $\{K(l);\ l \in \mathcal{L}_0\}$ be a finite polytopial complex of $\Delta(\mathcal{J})$. It is defined by two finite families $\left\{c_t \in \mathbb{R}^J,\, b_t \in \mathbb{R};\ t \in \mathcal{T}\right\}$ and $\{\mathcal{T}(l) \subset \mathcal{T};\ l \in \mathcal{L}\}$ such that:

$$K(l) = \{y \in \Delta(\mathcal{J});\ \langle y, c_t\rangle \leq b_t,\, \forall t \in \mathcal{T}(l) \subset \mathcal{T}\}\,, \quad \forall l \in \mathcal{L}_0.$$

Let us define $(c_{t,l}, b_{t,l}) = (c_t, b_t)$ if $t \in \mathcal{T}(l)$ and $(c_{t,l}, b_{t,l}) = (0,0)$ otherwise. Then we can rewrite $K(l) = \{y \in \Delta(\mathcal{J});\ \langle y, c_{t,l}\rangle \leq b_{t,l},\, \forall t \in \mathcal{T}\}$.

Definition 3.1 *A strategy σ is calibrated w.r.t. the complex $\{K(l);\ l \in \mathcal{L}_0\}$ if for every strategy τ of Nature, $\mathbb{P}_{\sigma,\tau}$-as:*

$$\limsup_{n\to\infty} \frac{|N_n(l)|}{n}\left(\langle \bar{\jmath}_n(l), c_{t,l}\rangle - b_{t,l}\right) \leq 0, \quad \forall t \in \mathcal{T}, \forall l \in \mathcal{L}_0.$$

Theorem 3.2 *There exist calibrated strategies w.r.t. any finite polytopial complex $\{K(l);\ l \in \mathcal{L}_0\}$.*

Proof. Consider the following auxiliary two-person game Γ'_c, where at stage $n \in \mathbb{N}$ the predictor (resp. Nature) chooses $l_n \in \mathcal{L}_0$ (resp. $j_n \in \mathcal{J}$) which generates the vector payoff $U_n \in \mathbb{R}^{TL_0}$ defined by:

$$U_n^{lk} = \begin{cases} \langle \mathbb{1}_{j_n=j}, c_{t,l} \rangle - b_{t,l} & \text{if } l = l_n \\ 0 & \text{otherwise.} \end{cases}$$

Any strategy that approaches the negative orthant Ω_- in Γ'_c is calibrated w.r.t. the complex $\{K(l);\ l \in \mathcal{L}_0\}$.

Blackwell's characterization of approachable convex sets (see Blackwell [5], Theorem 3) implies that the predictor can approach the convex set Ω_- if (and only if) for every mixed action of Nature in $\Delta(\mathcal{J})$, he has an action $x \in \Delta(\mathcal{L}_0)$ such that the expected payoff is in Ω_-. Given $y_n \in \Delta(\mathcal{J})$, choosing $l(y_n) \in \mathcal{L}_0$, where $l(y_n)$ is the index of the polytope that contains y_n, ensures that $\mathbb{E}_{y_n, l(y_n)}[U_n]$ is in Ω_-. Therefore there exist calibrated strategies with respect to any polytopial complex. □

This modification of the definition of calibration does not change the other part of our algorithms nor the remaining of the proofs (in particular, to calibrate the sequence of unobserved flags, the forecaster must use $\widehat{\gamma}_n$-perturbations). The constants in the rates of convergence are now smaller since L_0 can be much smaller than L and in Γ'_c, $\mathbb{E}[\|U_n\|^2]$ is bounded by $O\left(\frac{T_0}{\gamma_n}\right)$ where $T_0 = \sup_{l \in \mathcal{L}_0} T(l)$ is the maximum number of hyperplanes defining a polytope of the complex.

The main argument behind this algorithm (*i.e.* the characterization of approachable convex sets of Blackwell [5]) is quite close, in spirit, to the one of Lehrer & Solan [21]. Note that however, with our representation, the projection on Ω_- can be computed linearly in TL_0, so polynomially in L_0. Therefore, it reduces to the construction of an approachability strategy and so – as shown by Blackwell [5] – to the resolution, at each stage, of a linear programming of size polynomial in L_0.

3.2 Extension to the compact case

We prove in this section that the finiteness of $\mathcal{J}$ is not required.

Assume that instead of choosing j_n at stage $n \in \mathbb{N}$ – which generates the flag $f_n = \mathbf{s}(j_n)$ and an outcome vector $\left(\rho(i, j_n)\right)_{i \in \mathcal{I}}$ – Nature chooses directly an outcome vector $O_n \in [-1, 1]^I$ and a flag f_n which belongs to $\mathbf{s}(O_n)$ where $\mathbf{s}$ is a multivalued mapping from $[-1, 1]^I$ into $\Delta(\mathcal{S})^I$. As before, the decision maker's payoff is $O_n^{i_n}$ (the i_n-th coordinate of O_n) and he receives a signal

s_n whose law is $f_n^{i_n}$. Strategies of the forecaster and consistency are defined as before.

Theorem 3.3 *If the graph of* $\mathbf{s}$ *is a polytope, then there exists an internally consistent strategy* σ *such that, for every strategy* τ *of Nature, with* $\mathbb{P}_{\sigma,\tau}$ *probability at least* $1-\delta$*:*

$$\max_{l\in\mathcal{L}} \frac{|N_n(l)|}{n}\mathcal{R}_n(l) \leq O\left(\frac{1}{n^{1/3}}\sqrt{\ln\left(\frac{1}{\delta}\right)} + \frac{1}{n^{2/3}}\ln\left(\frac{1}{\delta}\right)\right). \tag{17}$$

The proof of this result is identical to the one of Theorem 2.10.

Note that the assumption that the graph of $\mathbf{s}$ is a polytope is fulfilled in the finite dimension case. The mapping $\mathbf{s}$ is multivalued since in finite dimension there might exist two different mixed actions y_1, y_1 in $\Delta(\mathcal{J})$ that generate the same outcome vectore (*i.e.* $\rho(\cdot, y_1) = \rho(\cdot, y_2) = O$) but different flags (*i.e.* $f_1 = \mathbf{s}(y_1) \neq \mathbf{s}(y_2) = f_2$). Hence we should have $f_1, f_2 \in \mathbf{s}(O)$.

3.3 Strengthening of the constants

We propose two different ideas to strengthen the constants of our algorithm. First, we can use (as did Lugosi, Mannor & Stoltz [23]) only one concentration inequality for every coordinate of the vector $U_{\omega,n}$ instead of one concentration inequality per coordinate. Second, we can implement sparser vector payoffs (so that its norm decreases) by looking at a slight different definition of calibration.

3.3.1 Concentration Inequalities in Hilbert Spaces

The rates of convergence of our algorithms rely mainly on three properties: Blackwell's approachability theorem, Hoeffding-Azuma's and Freedman's inequalities. These tools allowed us to study the convergence of a sequence of vectors $\bar{U}_n^+$ towards 0. Approachability is well defined for sequences of vectors, however the two concentration inequalities hold only for real valued martingales. To circumvent this issue, we used in the proofs the fact that if a process $\left\{U_n \in \mathbb{R}^d\right\}_{n\in\mathbb{N}}$ is a martingale then, for each coordinate, the process $\left\{U_n^k \in \mathbb{R}\right\}_{n\in\mathbb{N}}$ is a real valued martingale. This does not use the fact that U_n might be sparse and the use of concentration inequalities in Hilbert space can sharpen the constant.

Indeed, recall Hoeffding-Azuma's inequality:

Lemma 3.4 (Hoeffding[19], Azuma [3]) *Let U_n be a sequence of martingale differences bounded by K, i.e. for every $n \in \mathbb{N}$, $\mathbb{E}_{\sigma,\tau}[U_{n+1}|h_n] = 0$ and $|U_n| < K$.*

Then for every $n \in \mathbb{N}$ and every $\varepsilon > 0$:

$$\mathbb{P}_{\sigma,\tau}\left(\left|\bar{U}_n\right| \geq \varepsilon\right) \leq 2\exp\left(\frac{-n\varepsilon^2}{2K^2}\right),$$

which can be expressed as

$$\mathbb{P}_{\sigma,\tau}\left(\left|\bar{U}_n\right| \leq K\sqrt{\frac{2}{n}\ln\left(\frac{2}{\delta}\right)}\right) \geq 1-\delta. \tag{18}$$

Chen & White [11] proved an equivalent property for vector martingale in $\mathbb{R}^d$.

Lemma 3.5 (Chen & White [11]) *Let U_n be a sequence of martingale differences in $\mathbb{R}^d$ bounded almost-surely by $K > 0$. Then for every $n \in \mathbb{N}$ and for every $\varepsilon > 0$:*

$$\mathbb{P}_{\sigma,\tau}\left(\left\|\bar{U}_n\right\| \geq \varepsilon\right) \leq 2\max\left\{1, \sqrt{\frac{n\varepsilon^2}{2K^2}}\right\}\exp\left(\frac{-n\varepsilon^2}{2K^2}\right) \leq 2\exp\left(-\alpha\frac{n\varepsilon^2}{2K^2}\right),$$

for every $\alpha \leq 1 - \frac{1}{2e}$ (which equals approximatively 0.81).

Assume that for every $n \in \mathbb{N}$, $\|U_n\|_\infty \leq \|U\|_\infty$ and $\|U_n\|_2 \leq \|U\|_2$; we can deduce from the use of only Hoeffding-Azuma's inequality that:

$$\mathbb{P}_{\sigma,\tau}\left(\max_{l,k}\frac{|N_n(l)|}{n}\left|\bar{U}_n^{l,k}\right| \geq \varepsilon\right) \leq 2L^2\exp\left(\frac{-n\varepsilon^2}{2\|U\|_\infty^2}\right).$$

However, Chen and White's result, along with the fact that $\|U_n\| \leq L$, implies that:

$$\mathbb{P}_{\sigma,\tau}\left(\max_{l,k}\frac{|N_n(l)|}{n}\left|\bar{U}_n^{l,k}\right| \geq \varepsilon\right) \leq 2\exp\left(\frac{-n\varepsilon^2}{4\|U\|_2^2}\right)$$

which can reduce the dependency in L. The effects is even more dramatic when estimating the sequences of flags, since e_n has only positive component (so $\|e_n\|_\infty = \|e_n\|_2$).

There also exist variants of Bernstein's inequality (see *e.g.* Yurinskii [31]) in Hilbert spaces that can be used in order to get more precise constants.

3.3.2 Calibration with Respect of Neighborhoods

Definition 3.6 *Given a finite set* $\mathcal{Y} = \{y(l) \in \mathbb{R}^d, \omega(l) \in \mathbb{R};\ l \in \mathcal{L}\}$, $y(k)$ *is a neighbor of* $y(l)$ *if* $k \neq l$ *and the dimension of* $P(l) \cap P(k)$ *is equal to* $d-1$.

We defined a calibrated strategy with respect to $\mathcal{Y}$, as a strategy σ such that $\bar{j}_n(l)$ is asymptotically closer to $y(l)$ than to any other $y(k)$ as soon as the frequency of l does not go to zero. In fact, $\bar{j}_n(l)$ needs only to be closer to $y(l)$ than to any of its neighbors. So one can construct *neighbors*-calibrated strategies by modifying the algorithm given in Proposition 1.5; the payoff at stage n is now denoted by U'_n and is defined by:

$$\left(U'_n\right)^{lk} = \begin{cases} \|j_n - y(l)\|^2 - \|j_n - y(k)\|^2 & \text{if } l = l_n \text{ and } k \text{ is a neighbor of } l \\ 0 & \text{otherwise} \end{cases}$$

The strategy consisting in choosing an invariant measure of $\left(\bar{U}'_n\right)^+$ is calibrated and $M_n^2 = \sup_{m \leq n} \mathbb{E}_{\sigma,\tau}\left[\|U_m\|^2\right]$ equals $4\mathcal{N}$, where $\mathcal{N}$ is the maximal number of neighbors. This latter can be much smaller than 4, and the gain from this modification is limpid if we consider ε-calibration.

Indeed, in order to construct such strategies, we usually take any ε-discretization of $\Delta(J)$ so that $L = O\left(\varepsilon^{-(J-1)}\right)$. However, there exists a discretization such that $\mathcal{N} = 2^{-(J-1)}$, which is independent of ε.

A Proofs of technical results

This section is devoted to the proofs of previously mentioned results, *i.e.* Lemma 1.12 and Proposition 2.6.

A.1 Proof of Lemma 1.12

Let $l \in \mathcal{L}$ be fixed. we denote by $\mathcal{C} = \left\{c_t \in \mathbb{R}^d;\ t \in \mathcal{T}(l)\right\}$ the finite family of normal vectors to $(d-1)$-faces of $P(l)$ and by $\mathcal{B} = \{b_t \in \mathbb{R};\ t \in \mathcal{T}(l)\}$ the family of scalars such that :

$$P(l) = \left\{Z \in \mathbb{R}^d;\ \langle Z, c_t\rangle \leq b_t,\ \forall t \in \mathcal{T}(l)\right\}.$$

Any points satisfying Equation (5) belongs to

$$P_\varepsilon(l) = \left\{Z \in \mathbb{R}^d;\ \langle Z, c_t\rangle \leq b_t + \varepsilon,\ \forall t \in \mathcal{T}(l)\right\}.$$

For any vertex v of $P(l)$, there exists $t_1, \ldots, t_d \in \mathcal{T}(l)$ such that

$$v = \bigcap_{k=1}^{d} \Big\{ Z \in \mathbb{R}^d;\ \langle Z, c_{t_k} \rangle = b_{t_k} \Big\}$$

and $\{c_{t_1}, \ldots, c_{t_d}\}$ is a basis of $\mathbb{R}^d$. If we denote by v_ε the point defined by

$$v_\varepsilon = \bigcap_{k=1}^{d} \Big\{ Z \in \mathbb{R}^d;\ \langle Z, c_{t_k} \rangle = b_{t_k} + \varepsilon \Big\}$$

then $P_\varepsilon(l)$ is included in the convex hull of every v_ε.

Equation (5) can be rephrased as: if x belongs to $P_\varepsilon(l)$ then $d(x, P(l))$ is smaller than $M_P \varepsilon$. Therefore it is enough to prove this property for every v_ε since $d(\cdot, P(l))$ is a convex mapping thus maximized over a polytope on one of its vertices.

With these notations, for every $k \in \{1, \ldots, d\}$, $\langle v_\varepsilon - v, c_{t_k} \rangle = \varepsilon$ and there exists a unique decomposition $v_\varepsilon - v = \sum_{k=1}^{d} \alpha_k c_{t_k}$. Define the symmetric $d \times d$ Gram matrix Q_l by $Q_l^{kk'} = \langle c_{t_k}, c_{t_{k'}} \rangle$ and $\alpha = (\alpha_1, \ldots, \alpha_d)$. Then following classical properties hold:

1) $\|v_\varepsilon - v\|^2 = \alpha^T Q_l \alpha$ and there exist a $D = diag(\lambda_1, \ldots, \lambda_d)$ a diagonal matrix with $0 < \lambda_1 \leq \ldots \leq \lambda_d$ and a $d \times d$ matrix P and such that $P^{-1} = P^T$ and $Q_l = P^T D P$;

2) $Q\alpha = \underline{\varepsilon} = (\varepsilon, \ldots, \varepsilon)$ therefore $\alpha = Q_l^{-1} \underline{\varepsilon}$;

3) $\|v_\varepsilon - v\|^2 = (Q_l^{-1}\underline{\varepsilon})^T Q_l (Q_l^{-1}\underline{\varepsilon}) = \underline{\varepsilon}^T P^T D^{-1} P \underline{\varepsilon} \leq \varepsilon^2 d \lambda_1^{-1}$.

Therefore, for any $Z \in P_\varepsilon$ – and in particular for any point that satisfies Equation (5) –, $\|Z - \Pi_l(Z)\| \leq \max_v \|v_\varepsilon - v\| \leq \varepsilon . \sqrt{d} \sqrt{\lambda_1}^{-1}$. The result follows from the fact that L is finite. The constant M_P in Lemma 1.12 is smaller than the square root of the inverse of the smallest eigenvalue of all Q_l times $\sqrt{d}$; it depends on the inner products $\langle c_t, c_{t'} \rangle$ and on the dimension of $\mathcal{F}$.

A.2 Proof of proposition 2.6

Definition A.1 *Let K be a polytope. A correspondence $B : K \rightrightarrows \mathbb{R}^d$ is polytopial constant, if there exists $\{K(l);\ l \in \mathcal{L}\}$ a finite polytopial complex of K and $\{x(l);\ l \in \mathcal{L}\}$ such that $x(l) \in B(f)$ for every $f \in K(l)$.*

Let us now restate Proposition 2.6:

Proposition A.2 *BR is polytopial constant.*

This theorem is well-known and quite useful in the full monitoring case (see for example the Lemke-Howson [22] algorithm). In the *compact case*, Proposition 2.6 becomes:

Proposition A.3 *If* **s** *has a polytopial graph, then BR is polytopial constant.*

The proofs of both propositions rely on polytopial parameterized max-min programs defined in the next subsection.

A.2.1 Constant Solution of a Polytopial Parameterized Max-Min Program

A Polytopial Parameterized Max-Min Program (PPMP) is defined as follows. Let $\mathcal{X}$ and $\mathcal{Y}$ be two Euclidian spaces of respective dimension d_1 and d_2. Consider the program (P_f) - depending on a parameter f that belongs to some polytope $\mathcal{F}$ in $\mathbb{R}^{d_3}$ - that is defined by

$$(P_f): \quad \begin{array}{ccc} \max & \min & xAy, \\ x \in \mathcal{X} & y \in \mathcal{Y} & \\ s.t.\ Dx \leq d & s.t.\ E_f y \leq e_f & \end{array}$$

where A is a $d_1 \times d_2$ matrix, $\{E_f,\, e_f;\ f \in \mathcal{F}\}$ is a family of matrices and vectors (we do not specify the sizes the matrices, as long as each inequality makes sense) and D, d are also a fixed matrix and vector such that the admissible set $\mathcal{D} = \{x \in X;\ Dx \leq d\}$ is a polytope. The solution set of (P_f) is denoted by $B(f) \subset \mathcal{X}$ and this defines a multivalued mapping $B(\cdot)$ from $\mathcal{F}$ into $\mathcal{X}$.

Theorem A.4 *Assume that the correspondence S defined by:*

$$S: \begin{array}{ccc} \mathcal{F} & \rightrightarrows & \mathcal{Y} \\ f & \mapsto & S_f = \{y \in \mathcal{Y};\ E_f y \leq e_f\} \end{array}$$

has a polytopial graph **S**. *Then* $B : \mathcal{F} \rightrightarrows \mathcal{X}$ *is polytopial constant.*

Proof. Before going into full details, we first recall the following properties:

i) A linear program is minimized on a vertex of the polytopial feasible set (this is actually implied by the following point);

ii) Rockafella [26], Theorem 27.4, page 270: Given $x \in \mathcal{D}$ and $f \in \mathcal{F}$, if y minimizes xAy on S_f then

$$-xA \in NC_{S_f}(y),$$

where $NC_E(y)$ is the normal cone to the convex set $E \subset \mathbb{R}^d$ at $y \in E$ defined by :

$$NC_E(y) = \left\{ p \in \mathbb{R}^d;\ \langle p, z - y \rangle,\ \forall z \in E \right\};$$

iii) Ziegler [32], Example 7.3, page 193: If P is a polytope then the finite family $\{NC_P(v);\ v \text{ is a vertex of } P\}$ is a polyhedral complex of $\mathbb{R}^d$ called a normal fan (*i.e.* it is a finite family of polyhedra that cover $\mathbb{R}^d$ and such that each pair has an intersection with empty interior);

iv) Billera & Sturmfels [4], page 530: Since for every $f \in \mathcal{F}$, $S_f = \Pi^{-1}(f)$ where $\Pi : \mathbf{S} \subset \mathcal{F} \times \mathcal{Y} \to \mathcal{F}$ is the projection with respect to first coordinates, then there exists $\{K(l);\ l \in \mathcal{L}\}$, a polytopial complex of $\mathcal{F}$ such that the normal fan to S_f is constant on every $K(l)$ (this can alternatively be deduced from the following point);

v) Rambau & Ziegler [25], Proposition 2.4, page 221: On each of these polytopes $K(l)$, the mapping $f \mapsto S_f$ is linear. In particular, there exists a finite family of affine functions $Y(l)$ from $K(l)$ to $\mathcal{Y}$ such that the vertices of S_f are exactly $\{y(f);\ y(\cdot) \in Y(l)\}$.

Points i) and ii) imply that if x_f maximizes (P_f) – which is then minimized at some a vertex of S_f denoted by y_f, because of point i) – then it can be assumed that $-x_f A$ is a vertex of the polytope $NC_{S_f}(y_f) \cap \mathcal{D}_{A-}$ where $\mathcal{D}_{A-} := \{-xA;\ x \in \mathcal{D}\}$. Thus $B(f)$, the solution set to (P_f) contains at least an element of

$$\mathbf{X}_f = \left\{ x \in \mathcal{D}; -xA \text{ vertex of } \mathcal{D}_{A-} \cap NC_{S_f}(y_f), y_f \text{ vertex of } S_f \right\}.$$

By point iii), the normal fan and therefore $\mathbf{X}_f$ are constant on $K(l)$. The latter can also be assumed to be finite by taking a unique representant $x \in \mathbf{X}_f$ for every vertices of the intersection of the normal fan and $\mathcal{D}_{A-}$. Since the number of different fans is finite, for any $f \in \mathcal{F}$, the solution set to (P_f) contains at least an element of the finite set $\mathbf{X} = \bigcup_{f \in \mathcal{F}} \mathbf{X}_f$.

Moreover, for every $\mathbf{x} \in \mathbf{X}$:

$$\begin{aligned}
B^{-1}(\mathbf{x}) &= \left\{ f \in \mathcal{F};\ \min_{y \in S_f} \mathbf{x}Ay \geq \max_{x' \in \mathcal{D}} \min_{y \in S_f} x'Ay \right\} \\
&= \bigcup_{l \in \mathcal{L}} \left\{ f \in K(l);\ \min_{y \in S_f} \mathbf{x}Ay \geq \max_{x' \in \mathcal{D}} \min_{y \in S_f} x'Ay \right\} \\
&= \bigcup_{l \in \mathcal{L}} \bigcap_{\mathbf{x}' \in \mathbf{X}} \left\{ f \in K(l);\ \min_{y \in S_f} \mathbf{x}Ay \geq \min_{y \in S_f} \mathbf{x}'Ay \right\} \\
&= \bigcup_{l \in \mathcal{L}} \bigcap_{\mathbf{x}' \in \mathbf{X}} \bigcup_{y'(\cdot) \in Y(l)} \left\{ f \in K(l);\ \min_{y \in S_f} \mathbf{x}Ay \geq \mathbf{x}'Ay'(f) \right\} \\
&= \bigcup_{l \in \mathcal{L}} \bigcap_{\mathbf{x}' \in \mathbf{X}} \bigcup_{y'(\cdot) \in Y(l)} \bigcap_{y(\cdot) \in Y(l)} \left\{ f \in K(l);\ \mathbf{x}Ay(f) \geq \mathbf{x}'Ay'(f) \right\},
\end{aligned}$$

where, respectively, the second line is a consequence of point iv), the third line of the definition of $\mathbf{X}$ and the fourth and fifth lines of points i) and v).

By point v), the two mapping $y(\cdot)$ and $y'(\cdot)$ are affine on $K(l)$, so each possible set

$$\left\{ f \in K(l);\ \mathbf{x}Ay(f) \geq \mathbf{x}'Ay'(f) \right\}$$

is a polytope as the intersection of an half-space and the polytope $K(l)$. Since, the intersection of a union of polytopes remains a union of polytopes, for every $\mathbf{x} \in \mathbf{X}$, $B^{-1}(\mathbf{x})$ is a finite union of polytopes and B is polytopial constant. □

We can now prove simultaneously Propositions A.2 and A.3:

A.2.2 Proof of Propositions A.2 and A.3

Since $\mathbf{s}$ is linear, its graph, denoted by $\mathbf{S}$, is a polytope. Theorem A.4 (with $\mathcal{D} = \Delta(\mathcal{I})$) implies that the solution, denoted by $B(f)$ for every $f \in \mathcal{F}$, of the parameterized program

$$\max_{x \in \Delta(\mathcal{I})} \min_{y \in \mathbf{s}^{-1}(f)} \rho(x, y)$$

is polytopial constant. We denote by $\{K(l);\ l \in \mathcal{L}\}$ a corresponding polytopial complex. If B is constant on $K(l)$, then it is also constant on $\widehat{K}(l) = \Pi_{\mathbf{S}}^{-1}(K(l))$, which is a finite union of polytopes. □

Acknowledgements: I deeply thank my PhD advisor Sylvain Sorin for its great help and support. I also acknowledge very useful comments of Gilles Stoltz.

References

[1] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The non-stochastic multiarmed bandit problem. *SIAM J. Comput.*, 32:48–77 (electronic), 2002/03.

[2] F. Aurenhammer. A criterion for the affine equivalence of cell complexes in $\mathbb{R}^d$ and convex polyhedra in $\mathbb{R}^{d+1}$. *Discrete Comput. Geom.*, 2:49–64, 1987.

[3] K. Azuma. Weighted sums of certain dependent random variables. *Tôhoku Math. J. (2)*, 19:357–367, 1967.

[4] L. J. Billera and B. Sturmfels. Fiber polytopes. *The Annals of Mathematics*, 135(3):pp. 527–549, 1992.

[5] D. Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific J. Math.*, 6:1–8, 1956.

[6] D. Blackwell. Controlled random walks. In *Proceedings of the International Congress of Mathematicians, 1954, Amsterdam, vol. III*, pages 336–338, 1956.

[7] A. Blum and Y. Mansour. From external to internal regret. *J. Mach. Learn. Res.*, 8:1307–1324 (electronic), 2007.

[8] R. C. Buck. Partition of space. *Amer. Math. Monthly*, 50:541–544, 1943.

[9] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, Cambridge, 2006.

[10] N. Cesa-Bianchi, G. Lugosi, and G. Stoltz. Minimizing regret with label efficient prediction. *IEEE Trans. Inform. Theory*, 51:2152–2162, 2005.

[11] X. Chen and H. White. Laws of large numbers for Hilbert space-valued mixingales with applications. *Econometric Theory*, 12:284–304, 1996.

[12] A. P. Dawid. The well-calibrated Bayesian. *J. Amer. Statist. Assoc.*, 77:605–613, 1982.

[13] D. P. Foster and R. V. Vohra. Calibrated learning and correlated equilibrium. *Games Econom. Behav.*, 21:40–55, 1997.

[14] D. P. Foster and R. V. Vohra. Asymptotic calibration. *Biometrika*, 85:379–390, 1998.

[15] D. A. Freedman. On tail probabilities for martingales. *Ann. Probability*, 3:100–118, 1975.

[16] D. Fudenberg and D. K. Levine. Conditional universal consistency. *Games Econom. Behav.*, 29:104–130, 1999.

[17] J. Hannan. Approximation to Bayes risk in repeated play. In *Contributions to the Theory of Games*, volume 3 of *Annals of Mathematics Studies*, pages 97–139. Princeton University Press, Princeton, N. J., 1957.

[18] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

[19] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[20] T. Jaksch, R. Ortner, and P. Auer. Near-optimal regret bounds for reinforcement learning. *J. Mach. Learn. Res.*, 11:1563–1600, 2010.

[21] E. Lehrer and E. Solan. Learning to play partially-specified equilibrium. *manuscript*, 2007.

[22] C. E. Lemke and J. T. Howson, Jr. Equilibrium points of bimatrix games. *J. Soc. Indust. Appl. Math.*, 12:413–423, 1964.

[23] G. Lugosi, S. Mannor, and G. Stoltz. Strategies for prediction under imperfect monitoring. *Math. Oper. Res.*, 33:513–528, 2008.

[24] V. Perchet. Calibration and internal no-regret with random signals. *Proceedings of the 20th International Conference on Algorithmic Learning Theory*, pages 68–82, 2009.

[25] J. Rambau and G. M. Ziegler. Projections of polytopes and the generalized Baues conjecture. *Discrete Comput. Geom.*, 16:215–237, 1996.

[26] R. T. Rockafellar. *Convex Analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[27] A. Rustichini. Minimizing regret: the general case. *Games Econom. Behav.*, 29:224–243, 1999.

[28] E. Seneta. *Nonnegative Matrices and Markov Chains*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1981.

[29] S. Sorin. Supergames. In *Game theory and applications (Columbus, OH, 1987)*, Econom. Theory Econometrics Math. Econom., pages 46–63. Academic Press, San Diego, CA, 1990.

[30] S. Sorin. *Lectures on Dynamics in Games*. Unpublished Lecture Notes, 2008.

[31] V. Yurinskii. Exponential inequalities for sums of random vectors. *Journal of Multivariate Analysis*, 6:473 – 499, 1976.

[32] G. Ziegler. *Lectures on Polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.